

(19)日本国特許庁 (JP)

## (12)公開特許公報 (A)

(11)特許出願公開番号

特開2001-34292

(P2001-34292A)

(43)公開日 平成13年2月9日(2001.2.9)

(51)Int.C1.<sup>7</sup>  
G10L 15/28  
15/18

識別記号

F I  
G10L 3/00テーマコード (参考)  
H 5D015  
537 G

審査請求 未請求 請求項の数34 O L (全28頁)

(21)出願番号

特願平11-210819

(22)出願日

平成11年7月26日(1999.7.26)

(71)出願人 000004260  
株式会社デンソー

愛知県刈谷市昭和町1丁目1番地

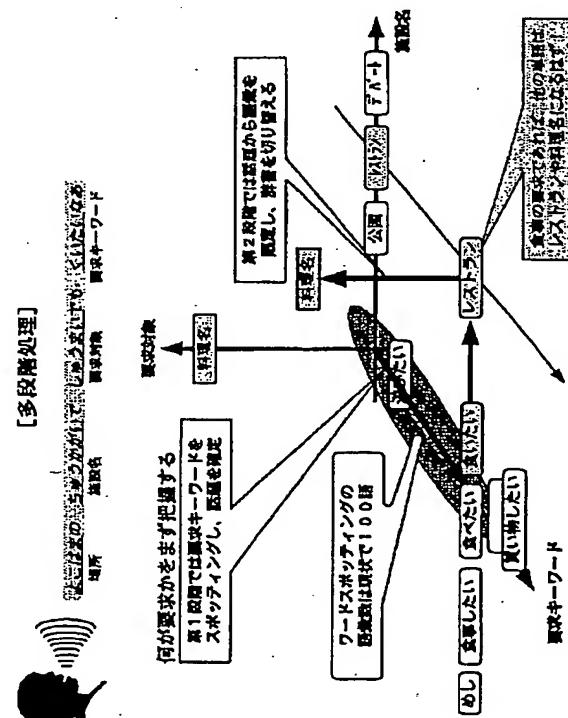
(72)発明者 笹木 美樹男  
愛知県刈谷市昭和町1丁目1番地 株式会  
社デンソー内(72)発明者 浅見 克志  
愛知県刈谷市昭和町1丁目1番地 株式会  
社デンソー内(74)代理人 100082500  
弁理士 足立 勉  
Fターム(参考) 5D015 AA04 BB01 HH13 HH23 LL02  
LL04 LL05 LL06

(54)【発明の名称】単語列認識装置

(57)【要約】

【課題】誤認識に起因し、利用者が途方に暮れてしまうといった不都合を防止可能な単語列認識装置を提供する。

【解決手段】多段階処理を説明する。発話単語と単語属性とが次のように分類されたとする。「よこはま(場所)のちゅうかがい(施設名)でしゅうまい(要求対象)でもくいたい(要求キーワード)なあ」。したがって、まず、何が要求であるかを把握するため、第1段階では「くいたい」という要求キーワードをスポットティングし、話題を確定する。そして、第2段階では、第1段階で確定させた話題から語彙を限定し、目的地をベースとしたレストラン名と関連する料理名で100語の大半を構成した辞書に切り替える。これは、「くいたい」という要求キーワードから食事の要求であることが判るため、単語列を構成する他の単語はレストラン名や料理名となっていると予想できるからである。



## 【特許請求の範囲】

【請求項1】認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い単語列の候補を出力する単語列出力手段と、

その単語列出力手段から単語列を構成する単語が順次出力される毎に、その出力単語の内容に基づいて次に入力を期待する語彙群を利用者が視認可能なように提示する漸進的階層探索手段とを備えることを特徴とする単語列認識装置。

【請求項2】請求項1記載の単語列認識装置において、前記漸進的階層探索手段は、前記単語列出力手段から1の単語が输出されてから所定時間以上次の単語が输出されなかった場合に限って、前記期待語彙群を提示することを特徴とする単語列認識装置。

【請求項3】請求項1又は2記載の単語列認識装置において、

前記漸進的階層探索手段は、前記単語列出力手段から出力された単語が、その直前に出力された単語に対応する前記期待語彙群以外の単語であった場合には、前記直前に出力された単語に対応する前記期待語彙群を再度提示することを特徴とする単語列認識装置。

【請求項4】認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い単語列の候補を出力する単語列出力手段と、

その単語列出力手段から出力された単語列を構成する単語の内、所定の属性を持つ単語に基づいて話題を確定し、統いて、その確定した話題に基づく語彙に限定した前記認識用辞書データを再構成し、その再構成した認識用辞書データに基づいて他の単語を認識するという処理を必要なだけ繰り返し実行する多段階処理手段とを備えることを特徴とする単語列認識装置。

【請求項5】請求項4記載の単語列認識装置において、前記多段階処理手段が最初に前記話題を確定するために扱う単語属性は、要求内容を示す属性であることを特徴とする単語列認識装置。

【請求項6】請求項4記載の単語列認識装置において、前記話題を確定するために扱う所定の属性を持つ単語が複数存在する場合は、構文位置に基づいていずれか1の単語を選択することを特徴とする単語列認識装置。

【請求項7】請求項4記載の単語列認識装置において、前記話題を確定するために扱う所定の属性を持つ単語は、所定レベル以上の尤度を備えている場合にのみ採用することを特徴とする単語列認識装置。

【請求項8】請求項4～7のいずれか記載の単語列認識装置において、

前記話題に基づく語彙に限定した認識用辞書データを再構成する際には、現在の話題に連関する話題に対応する語彙についても考慮することを特徴とする単語列認識装置。

【請求項9】請求項4～8のいずれか記載の単語列認識 50

装置において、

前記話題に基づく語彙に限定した認識用辞書データを再構成する際には、文脈の連続性という観点から定まる語彙についても考慮することを特徴とする単語列認識装置。

【請求項10】請求項4～9のいずれか記載の単語列認識装置において、

さらに、認識対象の要求を推定する要求推定手段を備え、

10 前記話題に基づく語彙に限定した認識用辞書データを再構成する際には、前記要求推定手段にて推定した要求という観点から定まる語彙についても考慮することを特徴とする単語列認識装置。

【請求項11】請求項4～10のいずれか記載の単語列認識装置において、

前記話題に基づく語彙に限定した認識用辞書データを再構成する際には、不連続的に発生する予測不可能な緊急事態や警告対象となる事態に対応する語彙として予め設定した語彙を含めることを特徴とする単語列認識装置。

20 【請求項12】認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い単語列の候補を出力する単語列出力手段と、その単語列出力手段から出力された単語列が文脈に沿った期待通りの内容であるか否かを判断し、期待から外れていると判断した場合は、少なくとも話題が転換されたのかどうかを確認するための問い合わせを行う話題転換確認処理、話題が転換されたことを宣言する話題転換宣言処理、それまでの話題が継続していると仮定して文脈に沿った対応を行う文脈優先対応処理のいずれかを行う期待外時対応手段とを備えることを特徴とする単語列認識装置。

【請求項13】請求項12記載の単語列認識装置において、

前記単語列出力手段は、前記単語列に付随して尤度も出力可能であり、

前記期待外時対応手段は、前記単語列に付随する尤度に基づいて、前記文脈に沿った期待通りの内容であるか否かを判断することを特徴とする単語列認識装置。

【請求項14】請求項13記載の単語列認識装置において、

前記期待外時対応手段は、単語列出力手段から出力された単語列の候補が1つであり、且つ前記単語列に付随する尤度が所定値以上であれば、前記話題転換確認処理を実行することを特徴とする単語列認識装置。

【請求項15】請求項12～14のいずれか記載の単語列認識装置において、

前記期待外時対応手段は、前記話題転換確認処理の実行後に前記単語列出力手段から出力された単語列が、その転換された話題に沿った内容であれば、前記話題転換宣言処理を実行することを特徴とする単語列認識装置。

【請求項16】請求項12～15のいずれか記載の単語列認識装置において、

前記期待外時対応手段は、前記話題が転換された場合であっても、転換直後には前記文脈優先対応処理を実行し、その後も前記転換された話題が継続している場合に限って、前記話題転換確認処理を実行することを特徴とする単語列認識装置。

【請求項17】認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い単語列の候補を出力する単語列出力手段と、

その単語列出力手段から出力された単語列が所定の否定的内容であるか否かを判断し、所定の否定的内容であると判断した場合は、所定の誤認識対応処理を実行する誤認識対応手段を備えることを特徴とする単語列認識装置。

【請求項18】請求項17記載の単語列認識装置において、

前記誤認識対応手段は、正しい話題が何であるかを確認するための問い合わせ返し処理を実行することを特徴とする単語列認識装置。

【請求項19】請求項17又は18記載の単語列認識装置において、

前記誤認識対応手段は、要求する内容に対応したキーワードを含める指示を利用者が視覚又は聴覚にて認識可能なように提示するヘルプモード処理を実行することを特徴とする単語列認識装置。

【請求項20】請求項17又は18記載の単語列認識装置において、

前記誤認識対応手段は、初期状態に戻る初期化処理を実行することを特徴とする単語列認識装置。

【請求項21】請求項19又は20記載の単語列認識装置において、

前記誤認識対応手段は、正しい話題が何であるかを確認するための問い合わせ返し処理を所定回数を超えた場合には、前記ヘルプモード処理又は初期化処理を実行することを特徴とする単語列認識装置。

【請求項22】認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い単語列の候補を出力する単語列出力手段と、

現在の話題に連関する話題に対応する語彙を考慮した上で、前記認識用辞書データを再構成する辞書構成手段を備えることを特徴とする単語列認識装置。

【請求項23】認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い単語列の候補を出力する単語列出力手段と、

文脈の連続性という観点から定まる語彙を考慮した上で、前記認識用辞書データを再構成する辞書構成手段を備えることを特徴とする単語列認識装置。

【請求項24】認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い

単語列の候補を出力する単語列出力手段と、認識対象の要求を推定する要求推定手段と、

前記要求推定手段にて推定した要求という観点から定まる語彙を考慮した上で、前記認識用辞書データを再構成する辞書構成手段とを備えることを特徴とする単語列認識装置。

【請求項25】認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い単語列の候補を出力する単語列出力手段と、

10 不連続的に発生する予測不可能な緊急事態や警告対象となる事態に対応する語彙として予め設定した語彙を含めて前記認識用辞書データを再構成する辞書構成手段とを備えることを特徴とする単語列認識装置。

【請求項26】請求項22記載の単語列認識装置において、

前記辞書構成手段は、認識対象のシステムで扱う現実の話題に対応して予め各カテゴリ毎の語彙クラスタに分割されて構成されるクラスタ辞書間の関係を考慮して前記認識用辞書データを再構成することを特徴とする単語列認識装置。

20 【請求項27】請求項22記載の単語列認識装置において、

前記辞書構成手段は、単語間の意味的関係を考慮して前記認識用辞書データを再構成することを特徴とする単語列認識装置。

【請求項28】請求項27記載の単語列認識装置において、

前記単語間の意味的関係とは単語間の包含関係であることを特徴とする単語列認識装置。

30 【請求項29】請求項27記載の単語列認識装置において、

前記単語間の意味的関係とは単語間の連想関係であることを特徴とする単語列認識装置。

【請求項30】請求項23記載の単語列認識装置において、

前記辞書構成手段は、発話に対する応答の妥当性を考慮して前記認識用辞書データを再構成することを特徴とする単語列認識装置。

40 【請求項31】請求項23記載の単語列認識装置において、

前記辞書構成手段は、発話とその発話に対する応答で構成される対話ユニット同士の関係の妥当性を考慮して認識用辞書データを再構成することを特徴とする単語列認識装置。

【請求項32】認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い単語列の候補を出力する単語列出力手段と、

前記請求項1～3のいずれかに記載した漸進的階層探索手段、前記請求項4～11のいずれかに記載した多段階処理手段、前記請求項12～16のいずれかに記載した

期待外時対応手段、前記請求項17～21のいずれかに記載した誤認識対応手段、前記請求項22～31のいずれかに記載して辞書構成手段の内の少なくとも2つの手段とを備えることを特徴とする単語列認識装置。

【請求項33】請求項1～32のいずれか記載の単語列認識装置において、

前記認識対象は人間であり、

前記単語列出力手段は、前記認識対象者によって入力された音声を認識用辞書データと比較した一致度合いの高い単語列の候補を出力することを特徴とする単語列認識装置。

【請求項34】請求項1～32のいずれか記載の単語列認識装置において、

前記認識対象は人間であり、

前記単語列出力手段は、前記認識対象者によって入力された手書き文字列を認識用辞書データと比較し、一致度合の高い複数の単語列候補を出力することを特徴とする単語列認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、例えば音声認識に代表される様々な信号の認識結果候補群が離散的な単語列の集合として得られる場合に、適正な単語列を認識することのできる単語列認識装置に関するものである。

【0002】

【従来の技術】従来より、例えば人間から発せられた音声を入力し、辞書データと比較して一致度合の高い複数の単語列候補を出力する認識装置が知られており、例えばその認識装置を用いて使用者の発話内容を認識し、認識結果に応じて機器を動作させる制御システムなどに適用されている。例えば、使用者が音声コマンドとして予め定められた言葉を発話すると、その言葉に対応した情報検索動作などを行う自動車用ナビゲーション装置などが実用化されている。

【0003】

【発明が解決しようとする課題】ところで、現状の音声認識には連続単語認識とワードスポットティングという2つの手法がある。前者の認識手法によれば、例えば「おかざき、〇〇〇〇〇（店名）、らーめん」といった単語毎の認識が可能であるが、実際には誤認識を避けて通ることはできない。そのため、現行のナビゲーション装置などでは、認識結果として複数の候補を持ち、最初にその内の1つをトーカバックしてユーザに確認を求めることがなされている。ユーザは自分が発話した内容と違つていれば、認識結果が違つてある旨を装置に伝える。すると、装置側は別の認識結果を提示して再度ユーザの確認を求める。認識結果の候補を多数準備して次々に提示していくれば、最終的にはユーザが意図した内容の認識結果となることは可能であるが、候補の中には、全く意味をなさない内容のものも含まれる可能性があり、上述し

た最終的に適切な候補に至るまでに長時間要してしまうことも考えられる。

【0004】また、後者のワードスポットティング手法は、例えば「えーと、おかざきの〇〇〇〇〇（店名）でらーめんくいたいなあ」という日常語的な音声入力から「おかざき、〇〇〇〇〇（店名）、らーめん、くいたい」というキーワードを抽出できるという点で近年急速に注目されている音声認識技術である。しかしながら、当該手法の場合には、ラティスと呼ばれるその出力（時区間情報と確率情報を持った単語集合）から生成される単語列候補の数は非常に多く、意味を持つ小数の単語列に絞られることは稀である。また、当該手法における現在の認識語彙数は100語程度であるが、将来的には1000語以上に増大することが見込まれ、その結果として発生するラティスから生成される単語列は膨大な数に上ると考えられる。したがって、上述した連続単語列認識手法の場合で述べたのと同様の（むしろ、より顕著な）問題が存在する。

【0005】また、現状の音声認識技術では、雑音や音声環境の変動などに基づく誤認識時において、正常な対話への復帰が困難であった。例えば地名を入力したい場合に、意図せず駅名中から選択するモードに入ってしまうと、それ以外の地名を適切に認識してもらえないくなる。誤認識を少なくするために絞り込みができる辞書構造にしておくことは、ある意味では好ましく、多くの単語認識手法において用いられているが、上述したように意図しない認識モードに陥ってしまうと、それ以降の所望の入力ができない可能性がある。そして、そのモードからエスケープする操作を知らないと正常な対話への復帰が困難となり、ユーザは途方に暮れてしまうこともありえる。

【0006】本発明は、こうした問題に鑑みなされたものであり、誤認識に起因し、例えば利用者の意図しない認識モードに陥って復帰ができず利用者が途方に暮れてしまう、といった不都合を防止可能な単語列認識装置を提供することを目的とする。

【0007】

【課題を解決するための手段、及び発明の効果】上記目的を達成するため、大きく分けて以下の4つの対応処理を提案する。それは、漸進的階層探索（請求項1～3）、多段階処理（請求項4～11）、期待外時対応（請求項12～16）、誤認識対応（請求項17～21）、辞書の動的構成（請求項22～31）である。以下、順番に説明する。

【0008】（1）漸進的階層探索

請求項1に記載の単語列認識装置は、単語列出力手段が、認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い単語列の候補を出力する。そして、漸進的階層探索手段が、その単語列出力手段から単語列を構成する単語が順次出力され

る毎に、その出力単語の内容に基づいて次に入力を期待する語彙群を利用者が視認可能なように提示する。

【0009】認識対象の動作内容が反映された情報としては、認識対象者から発話された音声による情報をはじめ、他にも種々考えられる。これらについては後述することにして、ここでは、この認識装置の利用者から発話された場合、つまり音声認識装置として実現した場合を例に挙げて説明することにする。

【0010】例えば「岡崎で食事したいなあ、インド料理がいいね」という発話が利用者からなされた場合、ワードスポットティングによる音声認識手法の出力結果は、「岡崎、食事、インド料理」という単語列になる。従来の音声認識手法では、これら3つが揃った段階で認識に対応するシステム側の処理が開始されていたが、本発明の漸進的階層探索手段によれば、「岡崎」が入力された時点で、次に利用者に入力を期待する発話語彙を即座に提示できるため、利用者はとまどうことなく発話できる。これによって、誤認識の原因となる認識辞書外の語彙を利用者が発話してしまうことを未然に防止できる。

【0011】なお、請求項2に示すように、漸進的階層探索手段は、単語列出力手段から1の単語が出力されながら所定時間以上次の単語が出力されなかった場合に限って期待語彙群を提示するようにしてよい。つまり、「提示できる状態にしておき」実際に提示するのは、所定時間以上次の単語が出力されなかった場合に限るのである。これは、「岡崎で食事したいなあ」のように次の単語が連続して発話されることが多いような状況では、むしろ提示しない方がよい場合もあるからである。そのため、上述例で言えば、岡崎の後に所定時間（例えば1～2秒）の無音区間があった場合に、利用者は次の語彙入力に迷っていると推定し発話語彙群を提示するのである。

【0012】また、請求項3に示すように、漸進的階層探索手段は、単語列出力手段から出力された単語が、その直前に出力された単語に対応する期待語彙群以外の単語であった場合には、直前に出力された単語に対応する期待語彙群を再度提示するようにしてよい。例えば「岡崎」に対応する期待語彙群として「食事、喫茶店、デパート、公園、駐車場、……」が準備されていた場合に、それらのいずれにも該当しなければ、誤認識であるとして、再度「岡崎」に対応する期待語彙群を提示する。このようにすれば、「岡崎」という場所の入力結果を保ったままで、認識作業を継続することができる。

### 【0013】(2) 多段階処理

請求項4記載の単語列認識装置は、多段階処理手段が、単語列出力手段から出力された単語列を構成する単語の内、所定の属性を持つ単語に基づいて話題を確定し、統いて、その確定した話題に基づく語彙に限定した認識用辞書データを再構成し、その再構成した認識用辞書データに基づいて他の単語を認識するという処理を必要など

け繰り返し実行する。

【0014】例えば、現状のワードスポットティング手法では1回の処理に対する認識語彙数は100語程度であり、連続単語認識の辞書のような大規模化は困難である。一方、実用的に見た場合、特定の話題（例えば「食事に行く」など）にフォーカスする際には、例えば100語程度でも対応可能である。したがって、話題を的確に認識し、誤認識の際は話題の不連続性を検出し、これらに応じてワードスポットティングの語彙を切り替えていければよい。そこで、多段階処理を行う。このようにすれば、利用者が途方に暮れてしまわないような単語列の認識を、小語彙で達成できるようになり、上述のワードスポットティング手法を採用した場合の不都合を解消できる。

【0015】ところで、多段階処理手段が最初に話題を確定するために扱う単語属性としては、例えば請求項5に示すように、要求内容を示す属性であることが考えられる。利用者の発話はたいていのコンテンツ検索の場合、『場所』『施設名』『要求対象』『要求キーワード』からなるか、その並び替え、あるいは省略形で基本形が構成されると考えられる。この内、一般的には何が要求であるかを把握することが好ましい。それは、要求を把握するために想定する要求キーワードが、場所や施設名などに比較して少ない数でもよいからである。なお、要求キーワード以外の『場所』、『施設名』又は『要求対象』を、最初に話題を確定するための単語属性として扱うことを否定するものではないことを付言しておく。

【0016】なお、1の単語列中に、話題を確定するために扱う所定の属性を持つ単語が複数存在する場合もある。その場合は、請求項6に示すように構文位置に基づいていはずれか1の単語を選択することが考えられる。例えば要求キーワードを探す場合であれば、構文全体として見た場合に、その構文における要求が何であるかを示す要求キーワードが存在する可能性の高い位置というものが決まつてくるからである。これは、自然言語として、例えば文法、意味、常識などを加味すれば自ずと定まる。

【0017】また、請求項7に示すように、話題を確定するために扱う所定の属性を持つ単語は、所定レベル以上の尤度を備えている場合にのみ採用することが好ましい。例えば、最初に話題を確定するための単語属性としては「要求キーワード」が好ましいことを述べたが、尤度が低い場合にあえて要求キーワードにこだわる必要もない。その場合には、他の単語属性で尤度が相対的に高いものがあれば、それを採用する方が好ましいこともある。

【0018】ところで、話題に基づく語彙に限定した認識用辞書データを再構成する上では、より適切な語彙を選ぶことが重要になってくる。そこで、以下に示すいく

つかの考慮点が挙げられる。まず、請求項8に示すように、現在の話題に連関する話題に対応する語彙についても考慮することが考えられる。その一例を挙げれば、ショッピングという話題に対して、駐車場（手段）、レストラン（付随する行動）、バーゲン（よくある付帯事象）などの語彙を含めるのである。

【0019】また、請求項9に示すように、文脈の連続性という観点から定まる語彙についても考慮してもよい。例えばある話題（ショッピングなど）で閉じた認識語彙の範囲で対話が継続すると考えて、このような話題の連続性という制約のもとで認識用辞書データの再構成をするのである。また、発話と応答の組合せを発話対あるいは対話ユニットと呼ぶこととすると、この対話ユニット内の妥当性という制約のものでの再構成、つまり発話に対して妥当な応答の範囲で認識用辞書データを再構成してもよい。さらには、対話ユニット間における接続性についても考慮することが考えられる。

【0020】一方、請求項10に示すように、認識対象の要求を推定する要求推定手段をさらに備え、その要求推定手段にて推定した要求という観点から定まる語彙についても考慮することが考えられる。これは、認識対象から直接入力（音声認識であれば利用者からの発話）されたものではなく、利用者の置かれている環境や状況あるいはユーザプロファイルなどを基に利用者の要求を推定し、利用者の次の発話として想定される内容に対応できるように認識用辞書データの再構成を行うのである。

【0021】また、請求項11に示すように、話題に基づく語彙に限定した認識用辞書データを再構成する際には、不連続的に発生する予測不可能な緊急事態や警告対象となる事態に対応する語彙として予め設定した語彙を含めてもよい。その一例として、単語列認識装置が車両に搭載されている場合を想定する。例えばカーナビゲーションシステムに組み込まれている場合などである。この際、突然の運転状況の変化などがここでいう「不連続的に発生する予測不可能な緊急事態や警告の対象となる事態」に相当する。例えば対話の最中に先行車が急停止したので急ブレーキをかけた場合、おもわず利用者が「あー、危なかった。」といってしまう状況は容易に考えられる。そのため、車両に搭載するのであれば、このような「危ない」という意味の語彙も含めておく。

#### 【0022】(3) 期待外時対応

請求項12に記載の単語列認識装置は、期待外時対応手段が、単語列出力手段から出力された単語列が文脈に沿った期待通りの内容であるか否かを判断し、期待から外れていると判断した場合は、少なくとも話題が転換されたのかどうかを確認するための問い合わせを行う話題転換確認処理、話題が転換されたことを宣言する話題転換宣言処理、それまでの話題が継続していると仮定して文脈に沿った対応を行う文脈優先対応処理のいずれかを行うのである。

【0023】なお、文脈に沿った期待通りの内容であるか否かを判断する場合には、請求項13に示すように、単語列に付随する尤度に基づくことが考えられる。そして、このように尤度に基づく場合には、期待外時対応手段が話題転換確認処理を実行する条件を次のように設定してもよい。つまり、請求項14に示すように、単語列出力手段から出力された単語列の候補が1つであり、且つ単語列に付随する尤度が所定値以上であることを条件にする。

【0024】また、期待外時対応手段が話題転換宣言処理を実行する条件は次のように設定することが考えられる。つまり、請求項15に示すように、話題転換確認処理の実行後に単語列出力手段から出力された単語列が、その転換された話題に沿った内容であることを条件にする。

【0025】また、文脈優先対応処理は、上述したように、それまでの話題が継続していると仮定して文脈に沿った対応を行うのであるが、話題の継続回数に基づいて話題の転換かどうかの判定を行い、それに応じて期待外時対応手段が話題転換確認処理と文脈優先対応処理とを使い分けてよい。つまり、請求項16に示すように、話題が転換された場合であっても、転換直後には文脈優先対応処理を実行し、その後も転換された話題が継続している場合に限って、話題転換確認処理を実行するのである。

【0026】このようにすることで、誤認識に起因して利用者が途方に暮れてしまう、といった不都合を防止することができる。

#### (4) 誤認識対応

請求項17に示すように単語列認識装置は、誤認識対応手段が、単語列出力手段から出力された単語列が所定の否定的内容であるか否かを判断し、所定の否定的内容であると判断した場合は、所定の誤認識対応処理を実行する。

【0027】ここで、誤認識対応処理としては、請求項18に示すように、正しい話題が何であるかを確認するための問い合わせ返し処理や、請求項19に示すように、要求する内容に対応したキーワードを含める指示を利用者が視覚又は聴覚にて認識可能なように提示するヘルプモード処理や、請求項20に示すように、初期状態に戻る初期化処理などであることが考えられる。

【0028】そして、請求項21に示すように、誤認識対応手段は、正しい話題が何であるかを確認するための問い合わせ返し処理を所定回数を超えた場合に、ヘルプモード処理又は初期化処理を実行することが考えられる。このような状況においては、そのままでは問題が解決されないことも考えられるので、単語列認識装置側が適切に認識できるようにキーワードを含めるように指示したり、あるいは初期状態に戻すことで、現状を打破するようにしている。なお、利用者の直接的な要求によっ

て、ヘルプモード処理や初期化処理を実行してもよい。

【0029】このようにすることで、誤認識に起因して利用者が途方に暮れてしまう、といった不都合を防止することができる。

#### (5) 辞書の動的構成

上述した(2)の多段階処理においては、単語列出力手段から出力された単語列を構成する単語の内、所定の属性を持つ単語に基づいて話題を確定し、続いて、その確定した話題に基づく語彙に限定した認識用辞書データを再構成し、その再構成した認識用辞書データに基づいて他の単語を認識するという処理を必要なだけ繰り返し実行した。しかし、このような多段階処理を前提としなくても、認識用辞書データの再構成をすることは、それだけでも効果はある。そこで、次のような辞書構成手段を備えるようにすることが考えられる。

【0030】まず、請求項22に示す単語列認識装置における辞書構成手段は、現在の話題に連関する話題に対応する語彙を考慮した上で認識用辞書データを再構成する。また、請求項23の場合の辞書構成手段は、文脈の連続性という観点から定まる語彙を考慮した上で、認識用辞書データを再構成する。

【0031】また、請求項24の場合は、認識対象の要求を推定する要求推定手段を備え、辞書構成手段が、その要求推定手段にて推定した要求という観点から定まる語彙を考慮した上で、認識用辞書データを再構成する。さらに、請求項25の場合は、辞書構成手段が、不連続的に発生する予測不可能な緊急事態や警告対象となる事態に対応する語彙として予め設定した語彙を含めて認識用辞書データを再構成する。

【0032】なお、これら請求項22～25に示した考慮内容の具体例については、請求項8～11において既に説明しているのでここでは繰り返さない。請求項8～11では多段階処理を前提とした内容であるのに対して、これら請求項22～25では多段階処理を前提とはしないが、このような辞書の再構成のみを実行することによっても、適切な語彙を選ぶことができる。

【0033】そして、このような辞書の再構成をするにあたっては、次の点を考慮することが好ましい。例えば、請求項26に示すように、認識対象のシステムで扱う現実の話題に対応して予め各カテゴリ毎の語彙クラスタに分割されて構成されるクラスタ辞書間の関係を考慮して認識用辞書データを再構成する。例えば施設→スキーチーム→○○○○スキーチーム（具体的な名称）といった関係が得られる。

【0034】また、請求項27に示すように、単語間の意味的関係を考慮して認識用辞書データを再構成してもよい。この場合には、例えば請求項28に示すように、単語間の意味的関係として単語間の包含関係を考慮したり、請求項29に示すように、単語間の意味的関係として単語間の連想関係を考慮することが考えられる。

【0035】単語間の包含関係としては、例えばスポーツに対して、テニス、スキー、スイミング、ジョギングなどが挙げられる。また、単語間の連想関係としては、同一分類のオブジェクトを連想する場合（例えば、うどんに対して麺類+ラーメン）や、シーンの構成要素を連想する場合（例えば、ゲレンデに対してスキー、スキーに対してリフト、スノーボード、ゴーグルなど）や、シーンに関連する興味の対象を連想する場合（例えば、スキーに対してゲレンデ、雪質、リフトなど）や、季節から代表的なシーンを連想する場合（例えば夏に対してプール、海水浴、かき氷、セミ、クーラーなど）や、要求キーワード間に基づく連想（例えば、はらへったに対してレストランなど）などが挙げられる。

【0036】一方、文脈の連続性という観点から定まる語彙を考慮した上で、認識用辞書データを再構成する場合には、請求項30に示すように、発話に対する応答の妥当性を考慮して認識用辞書データを再構成したり、請求項31に示すように、発話とその発話に対する応答で構成される対話ユニット同士の関係の妥当性を考慮して認識用辞書データを再構成することが考えられる。

【0037】ところで、これまでの説明では、漸進的階層探索、多段階処理、期待外時対応、誤認識対応、辞書の動的構成のいずれかを実行する単語列認識装置として説明したが、請求項32に示すように、認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較した一致度合いの高い単語列の候補を出力する単語列出力手段に加えて、請求項1～3のいずれかに記載した漸進的階層探索手段、請求項4～11のいずれかに記載した多段階処理手段、請求項12～16のいずれかに記載した期待外時対応手段、請求項17～21のいずれかに記載した誤認識対応手段、請求項22～31のいずれかに記載した辞書の動的構成手段の少なくとも2つを備える構成を採用しても良い。これらは、基本的には異なる原因で生じる誤認識を防止したり、あるいは誤認識が発生した後の対処であり、兼ね備えることによって、より適切な単語列の認識ができ、誤認識に起因して利用者が途方に暮れてしまう、といった不都合を防止することができる。

【0038】ところで、単語列認識装置に関しては、単語列出力手段が、認識対象の動作内容が反映された情報を入力し、認識用辞書データと比較して一致度合の高い単語列を出力するものであるという説明をしたが、具体的には、例えば次に示すようなものが考えられる。

【0039】まず、請求項33に示すように、認識対象者によって入力された音声を辞書データと比較し、一致度合の高い複数の単語列候補を出力する音声認識装置として実現することが考えられる。例えばカーナビゲーションシステムにおける目的地などの指示を音声入力するために実用化されており、適用対象としては一般的に考えられる。つまり、音声認識に際しては入力音声の音響

的な特徴に基づくのであるが、人間が発する音声の音響的な特徴は個人差が大きく、また日常的には正確な発音をしない場合も多いので誤認識が発生し易い。

【0040】また、例えば請求項34に示すように、認識対象者によって入力された手書き文字列を認識用辞書データと比較し、一致度合いの高い単語列を出力する文字認識装置として実現してもよい。手書き文字に関しても、上述の音響的特徴と同様に個人差が大きく、また日常的には正確な書体にて文字を書かない場合も多いので、やはり誤認識が発生し易い。

【0041】したがって、このような誤認識を防止したり、あるいは誤認識が発生した後の適切な対処を行うことによって、誤認識に起因して利用者が途方に暮れてしまう、といった不都合を防止することができる。なお、文字認識の形態については、種々考えられ、例えば筆記具で書いた文字をスキャナで読み取る場合の認識はもちろん、PDA（携帯情報端末）などによく見られるように、入力ペンにて画面上をなぞるような文字入力方法の場合の認識であってもよい。さらには、このような音声認識や文字認識のように、認識装置に入力される時点で直接的に単語列の内容となっているものに限らず、画像認識装置であってもよい。即ち、認識対象を捉えた画像を場面として認識した上で、場面を自然言語化するための辞書データと認識場面を比較し、一致度合いの高い複数の単語列候補を出力するような画像認識装置である。その一具体例として、認識対象を人間とし、その認識対象者が手話をしている画像から手話パターンを認識し、その手話パターンが表す自然言語的な意味を示す単語列を出力することが考えられる。この場合であれば、手話パターンと単語との対応パターンが確立されているので、そのパターンマッチングにより自然言語的な意味を示す単語列を出力することは容易に実現できる。但し、この場合の手話パターンについても、微妙な指使いによって表す単語が異なるため、手話をする者の個人差などによって、やはり誤認識は発生する。

【0042】したがって、やはりこの場合も、誤認識が含まれる可能性の高い手話パターンの認識装置においても、本発明を適用することで誤認識を防止したり、あるいは誤認識が発生した後の適切な対処を行うことができ、誤認識に起因して利用者が途方に暮れてしまう、といった不都合を防止することができる。

【0043】また、上述した単語列認識装置における漸進的階層探索手段、多段階処理手段、期待外時対応手段、誤認識対応手段、辞書構成手段の実行する処理をコンピュータシステムにて実現する機能は、例えば、コンピュータシステム側で起動するプログラムとして備えることができる。このようなプログラムの場合、例えば、フロッピーディスク、光磁気ディスク、CD-ROM、ハードディスク等のコンピュータ読み取り可能な記録媒体に記録し、必要に応じてコンピュータシステムにロー

ドして起動することにより用いることができる。この他、ROMやバックアップRAMをコンピュータ読み取り可能な記録媒体として前記プログラムを記録しておき、このROMあるいはバックアップRAMをコンピュータシステムに組み込んで用いても良い。

【0044】

【発明の実施の形態】以下、本発明の実施形態について、図面を用いて説明する。まず図1は、実施形態の単語列認識装置について機能に着目して概念的に示したブロック図である。

【0045】ユーザインターフェースを介して入力された、音声に代表される信号は、音声認識部あるいはその他の信号系認識部において所定の認識処理がなされ、単語列候補を出力する。音声認識部は音声認識及び単語列生成の機能を備えており、認識辞書（認識語彙辞書）を用いて、ある発話に相当する音声入力に対する認識を行い、認識辞書に登録されている単語の集まりとして単語列候補を得る。これがここでいう音声認識機能であるが、この時点では候補となる単語が同一時刻に複数個示される可能性がある。その後、単語列生成機能を用いることで、音声認識機能によって得た単語に基づき時区間に重なりのない単語列を複数個生成し、対話管理手段へ出力する。この際、単語列に付随して尤度も出力される。このように、1つの音声発話に対する音声認識結果は、通常、尤度情報を伴った候補が複数個存在することとなり、音声認識においては「湧き出し」と呼ばれる。

【0046】そして、この尤度情報を伴った単語列候補について、対話管理手段は文脈や意味的制約情報を適用して認識し、辞書構成手段を用いて、認識辞書を動的に更新する。さらに、対話管理手段は、画面制御手段を介して表示系に対する表示制御を行ったり、音声出力系、機器制御系に対する制御を行う。

【0047】以上は、機能に着目して単語列認識装置を概念的に示したブロック図であったが、次に、単語列認識装置を車載の制御システムに適用した場合の構成を示すブロック図である図2を参照して説明する。本制御システムは、自動車（車両）に搭載されて、ユーザとしての車両の乗員（主に、運転者）と音声にて対話しながら、その車両に搭載された様々な機器を制御するシステムである。

【0048】図2に示すように、本実施形態の制御システムは、制御装置1と、使用者が各種の指令やデータなどを外部操作によって入力するための入力装置3と、音声を入力するためのマイクロフォン5と、音声を出力するためのスピーカ7と、画像を表示するためのディスプレイ（表示装置）8と、車両の現在位置（現在地）の検出や経路案内などを行う周知のナビゲーション装置9と、車内の空調を制御するエアコン装置13と、カセットテープレコーダ、CD（コンパクトディスク）プレーヤ、MD（ミニディスク）プレーヤ、ラジオ、及びテレ

ビなどからなるオーディオ装置15と、周知のVICS (Vehicle Information and Communication System) の放送端末や、インターネットとの接続窓口であるインターネット放送端末との間で無線によりデータ通信を行う通信装置17と、車速や加減速状態などの車両運転状態、車両内外の温度、及び雨滴の有無などを検出するための各種センサ19と、車両のドアロック、窓ガラス(パワーウィンドウ)、エンジン、及びブレーキ装置などを制御する他の制御装置(図示省略)とが接続されて構成されている。

【0049】尚、ナビゲーション装置9は、車両の現在位置を検出するための周知のGPS装置や、地図データ、地名データ、施設名データなどの経路案内用データを記憶したCD-ROM、そのCD-ROMからデータを読み出すためのCD-ROMドライブ、及び、使用者が指令を入力するための操作キーなどを備えている。そして、ナビゲーション装置9は、例えば、使用者から操作キーを介して、目的地と目的地までの経路案内を指示する指令とが入力されると、車両の現在位置と目的地へ至るに最適な経路とを含む道路地図を、ディスプレイ8に表示させて経路案内を行う。また、ディスプレイ8には、ナビゲーション装置9によって経路案内用の道路地図が表示されるだけでなく、情報検索用メニューなどの様々な画像が表示され、さらに、オーディオ装置15がテレビのモードに設定されると、そのオーディオ装置15に備えられたテレビチューナにより受信されたテレビの受信画像が表示される。

【0050】そして、制御装置1は、CPU、ROM、及びRAMなどからなるマイクロコンピュータを中心に構成されたシステム制御部21と、システム制御部21に入力装置3からの指令やデータを入力するインターフェース(I/F)23と、マイクロフォン5から入力された音声信号をデジタルデータに変換してシステム制御部21に入力する音声入力部25と、システム制御部21から出力されたテキストデータをアナログの音声信号に変換してスピーカ7に出力し、スピーカ7を鳴動させる音声合成部27と、ディスプレイ8への表示画面の制御を行う画面制御部28と、上記ナビゲーション装置9、エアコン装置13、オーディオ装置15、通信装置17、各種センサ19、及び他の制御装置とシステム制御部21とをデータ通信可能に接続する機器制御インターフェース(機器制御I/F)29とを備えている。

【0051】また、制御装置1には、通信装置17によりインターネットから所望の情報を検索及び取得するために、インターネットのアドレス(インターネットアドレス)を記憶するインターネットアドレスデータベース31と、検索制御部33とが備えられている。そして、システム制御部21が、検索制御部33へ検索内容(コンテンツ)を表す検索キーワードを出力すると、検索制御部33は、機器制御I/F29を介し通信装置17を

動作させて、インターネット放送端末から上記検索キーワードに対応した情報を検索し、その検索結果をシステム制御部21へ入力させる。また、インターネットアドレスデータベース31には、検索制御部33によって過去に用いられたインターネットアドレスが、システム制御部21からの指令によって記憶され、検索制御部33は、システム制御部21から過去に入力した検索キーワードと同じ検索キーワードを受けると、インターネットアドレスデータベース31内のインターネットアドレスを再利用する。

【0052】一方また、制御装置1は、マイクロフォン5及び音声入力部25を介して入力される音声信号から、使用者が発話した言葉としてのキーワード(以下、発話キーワードともいう)を認識して取得するために、使用者が発話すると想定され且つ当該制御装置1が認識すべき複数の認識語彙を予め記憶した認識語彙記憶部34を備えている。つまり、この認識語彙記憶部34に記憶されている認識語彙群が、当該制御装置1の認識語彙データベースとなっている。

【0053】またさらに、制御装置1は、スピーカ7から出力する発話(以下、エージェント発話ともいう)の内容(即ち、スピーカ7の動作内容)とスピーカ7以外の他の機器Mの動作内容とを設定するため、及び、使用者の要求と使用者の精神的或いは肉体的な状態とを推定するためのデータを記憶する手段として、対話データベースを記憶する対話データ記憶部35と、要求・状態推定用データを記憶する要求・状態記憶部36と、使用者の複数人分の個人情報(以下、ユーザプロファイルともいう)を記憶するユーザプロファイル記憶部37とを備えている。尚、この対話データ記憶部35、要求・状態推定用データ記憶部36、及びユーザプロファイル記憶部37と、前述したインターネットアドレスデータベース31は、データの読み出しと書き込みとが可能な不揮発性メモリによって構成されている。

【0054】なお、対話データ記憶部35に記憶される対話データベース、要求・状態推定用データ記憶部36に記憶される要求・状態推定用データ、ユーザプロファイル記憶部37に記憶されるユーザプロファイルなどについては、例えば特願平10-162457号、特願平10-184840号などを参照されたい。

【0055】次に、以上のように構成された本実施形態の制御システムにおいて、システム制御部21で実行される処理の概要を図3を用いて説明する。なお、ここでは、図1で言えば音声認識部にて認識処理を行い、対話管理手段がその認識結果である単語列候補に対して所定の処理を行う「単語列の認識」にかかる処理について説明する。

【0056】処理が開始されると、入力された音声に対して認識処理を行う(S10)。そして、S20での判断結果に基づき、「入力待ち」とあると判断してS10

へ戻るか、多段階処理 (S 3 0) を行うか、認識完了又はタイムアウトであると判断して S 4 0 へ移行する。S 3 0 での多段階処理の詳細については、後述する。

【0057】S 4 0 での対話管理に移行した後は、期待外の応答であるかどうかを判断し (S 5 0) 、期待外の応答であれば S 6 0 へ移行し、期待外の応答の種類を判別する。その判別結果に応じて、話題転換の確認 (S 7 0) 、話題転換後の発話 (S 8 0) 、文脈優先の発話 (S 9 0) のいずれかを実行する。その後、発話処理 (S 1 8 0) を経て、S 1 0 へ戻る。

【0058】一方、S 5 0 にて期待外の応答ではない (つまり期待に沿った応答である) と判断された場合は、S 1 0 0 へ移行する。S 1 0 0 での判断処理において、エージェントからの問い合わせ回数が N 回を超えたと判断された場合は、ヘルプモード処理を実行する (S 1 2 0) 、また、S 1 0 0 での判断処理において、エージェントからの問い合わせ回数が K 回 (K > N) を超えた場合、又はユーザからのリセット要求があった場合は、S 1 7 0 へ移行して 初期状態に戻る。そして、これら以外の場合、すなわち、エージェントからの問い合わせ回数が N 回以下の場合には、S 1 1 0 での判断結果に基づき、ディスプレイ 8 に選択肢を提示する処理 (S 1 3 0) 、通常の発話戦略 (S 1 4 0) 、音声メニュー モード (S 1 5 0) 又は問い合わせ (S 1 6 0) のいずれかの処理を選択的に実行する。これらの詳細については後述する。

【0059】S 1 2 0 ~ S 1 7 0 のいずれかの処理が実行された後は、発話処理 (S 1 8 0) を経て、S 1 0 へ戻る。以上は、処理の概略的な流れの説明であったので、統いて詳細内容を説明していく。但しここでは、漸進的階層探索、多段階処理、期待外時対応処理、誤認識対応処理という本発明の概念単位の順番に説明することにする。但し、説明の都合上、期待外時対応処理と誤認識対応処理については、【3. 誤認識対応処理について】としてまとめ、その中で区別して説明している。

#### 【1. 漸進的階層探索について】

【1. 1 概要】漸進的階層探索は、図 3 のフローチャート中では、S 1 3 0 の選択肢を提示する処理に相当する。但し、1 単語ずつ処理するために他の処理とは時間サイクルが異なるため、事前に漸進的階層探索を実行するモードに設定しておく必要がある。したがって、漸進的階層探索モードに設定されている場合に限り、図 3 の S 1 3 0 は実行されることとなる。

【0060】【1. 2 具体例】図 4 には漸進的階層探索の具体的な画面遷移例を示す。

(1) 図 4 の最初の画面 G 1 では、デフォルトモードにおいて東海 4 県の県名が表示されている状態において「愛知県」と発話した結果、愛知県にフォーカスされたことを示している。

【0061】そして、制御システム側は、愛知県という

10

20

30

40

50

単語を認識した時点で「漸進的階層探索」機能を発揮して、次にユーザに期待する発話語彙を即座に画面に提示する。この場合は、画面 G 2 のように愛知県内の市町村名を表示する。なお、画面 G 2 では 4 つの市町村名しか挙げていないが、これは説明を簡単にするためのものである。なお、画面 G 2 は、利用者が「岡崎」と発話した結果、岡崎にフォーカスされたことを示している。

【0062】そして、制御システム側は、岡崎が入力された時点での「漸進的階層探索」機能を発揮して、次にユーザに期待する発話語彙を画面に提示できる状態にしておく。ここで「提示できる状態にしておく」としたのは、実際には、「岡崎で食事したいなあ」のように次の単語が連続して発話されることが多いので、実用上は提示しないからである。提示するのは、岡崎の後に所定時間 (例えば 1 ~ 2 秒) の無音区間があった場合には、ユーザが迷っていると推定し、発話語彙を画面提示する。つまり「岡崎の要求メニュー」である。その内容は、誤認識時に用いる画面 G 7 の岡崎の要求メニュー内の「もう一度お話下さい」を除いた部分となる。画面 G 3 が表示された状態で「インド料理がいいね。」と発話されると、画面 G 4 に示すようにインド料理にフォーカスされ、画面 G 5 に示すように、「愛知県岡崎市インド料理 検索しています」という表示し、検索が終了すると、画面 G 6 に示すように、その検索結果を表示する。

【0063】一方、画面 G 2 に示す岡崎が発話された時点で次に発話を期待する語彙以外の語彙が発話された場合には、画面 G 7 へ移行して「岡崎の要求メニュー」と共に「もう一度お話下さい」という案内を加えた内容を表示する。ここで、「食事」と発話されれば画面 G 3 へ移行し、「デパートは?」と発話されれば、画面 G 8 に示すように、該当するデパートを一覧表示する。そして、その内のいずれかが指定されて「△△△の地図」と発話されると、画面 G 9 に示すように、そのデパートの位置が明確になるように、周辺の地図と共に表示する。

【0064】(2) 図 5 も漸進的階層探索の一例である。図 5 の画面 G 1 ~ G 3 は図 4 にて示した画面内容と同じであるが、デフォルトモードにおいて東海 4 県の県名が表示されている状態において「東京」と発話すると、画面 G 1 1 へ移行する。この場合、東京といつても東京都のみを指すのではない場合もあるので、東京都周辺を対象としてもよい。

【0065】そして、制御システム側は、東京という単語を認識した時点で「漸進的階層探索」機能を発揮して、次にユーザに期待する発話語彙を即座に画面に提示する。この場合は、画面 G 1 2 に示すように東京都内の市区町村名を画面表示する。そして、銀座と発話した時点で銀座にフォーカスする。

【0066】そして、制御システム側は、銀座という単語を認識した時点で「漸進的階層探索」機能を発揮して、次にユーザに期待する発話語彙を画面に提示できる

状態にしておく。ここで「提示できる状態にしておく」としたのは、上述の画面G2→G3へ移行する部分と同様に、実際には、「銀座、〇〇〇ビル」のように次の単語が連続して発話されることが多いので、実用上は提示しないからである。提示するのは、銀座の後に所定時間（例えば1~2秒）の無音区間があった場合である。画面に提示する発話語彙は、銀座内の地名関連情報である。例えば地名そのものでもよいし、使い勝手の面から言えば、〇〇〇ビルや□□デパートのような施設名でもよい。

【0067】そして、利用者から例えば〇〇ビルと発話されると、画面G13に示すように、その〇〇〇ビルの位置が明確になるように、周辺の地図と共に表示する。

【1.3 効果】例えば「岡崎で食事したいなあ、インド料理がいいね」という発話が利用者からなされた場合、ワードスポットティングによる音声認識手法の出力結果は、「岡崎、食事、インド料理」という単語列になる。従来の音声認識手法では、これら3つが揃った段階で認識に対応するシステム側の処理が開示されていたが、本手法によれば、「岡崎」が入力された時点で、次

よこはまの ちゅうかがいで  
(場所) (施設名)

したがって、まず、何が要求であるかを把握するため、第1段階では「くいたい」という要求キーワードをスポットティングし、話題を確定する。

【0069】そして、第2段階では、第1段階で確定させた話題から語彙を限定し、辞書を切り替える。すなわち、この場合には、目的地をベースとしたレストラン名と関連する料理名で100語の大半を構成する。これは、「くいたい」という要求キーワードから食事の要求であることが判るため、単語列を構成する他の単語はレストラン名や料理名となっていると予想できるからである。これにより、「ちゅうかがい」や「しゅうまい」などが認識語彙としてヒットし易い辞書を構成することができる。

【0070】(2) なお、図6では第1段階の処理として要求キーワードをスポットティングして話題を確定しているが、それ以外の施設名や場所、あるいは要求対象をスポットティングして話題を確定してもよい。但し、現実的には、ワードスポットティングの語彙数は現状では100語程度であるので、その程度の語彙でまかなかうことを鑑みると、要求キーワードでの話題確定が好ましい。

【0071】(3) また、図6で示した具体例は、1の単語列を構成する単語の属性という観点からn次元の軸を設定したが、さらに時間軸に沿った関連性を考慮しても良い。つまり、文脈という観点も加味して話題を確定するのである。

【2.2 ユーザ発話の基本構成】ユーザの発話はたいていのコンテンツ検索の場合、『場所』『施設名』『要求対象』『要求キーワード』からなるか、その並び替

に利用者に入力を期待する発話語彙を即座に提示できるため、利用者はとまどうことなく発話できるようになる。これによって、誤認識の原因となる認識辞書外の語彙を利用者が発話してしまうことを未然に防止できる。

## [2. 多段階処理について]

### [2.1 概要]

(1) 現状のワードスポットティング手法では1回の処理に対する認識語彙数は100語程度であり、連続単語認識の辞書のような大規模化は困難である。一方、実用的

10 に見た場合、車室内において発生すると想定される特定の話題（例えば「食事に行く」など）にフォーカスする際には、例えば100語程度でも対応可能である。したがって、話題を的確に認識し、誤認識の際は話題の不連続性を検出し、これらに応じてワードスポットティングの語彙を切り替えていけばよい。そこで、多段階処理を行う。

【0068】図6には、「よこはまのちゅうかがいでしゅうまいでもくいたいなあ」という発話がなされた場合に行なう多段階処理の一例を示した。なお、本発話例で

20 は、以下のようないくつかの属性に分類できるものとする。

しゅうまいでも くいたいなあ  
(要求対象) (要求キーワード)

え、あるいは省略形で基本形が構成されると考えられる。語順が変わった場合には要求キーワードが音声信号中のどこに存在するかは不明であるが、例えば本願出願人が特願平11-20349号にて提案したような適正単語列の推定手法を用いることにより、構文的な制約に基づいて複数の候補に対して優先順序を定めることはできる。

30 【0072】[2.3 認識辞書の構成] 認識辞書は図2に示す認識語彙記憶部34に記憶されている認識語彙データベースから動的に構成し得るものとする。認識語彙データベースは、システムで扱う現実の話題に対応して予め各カテゴリ毎の語彙クラスタに分割しておく（図7参照）。なお、この各カテゴリ毎の語彙クラスタはクラスタ辞書と呼ばれる。また、カテゴリには例えば下記のような種類がある。

【0073】(1) 各種コマンド

・ナビコマンド

40 ④ スケジュール帳

・アドレス帳

・電話

(2) 要求キーワード (要求KW)

(3) 施設名

① レストラン名

・料理名

・雰囲気

・値段

② スキー場名

50 ③ ゴルフ場名

- ④デパート名
- ⑤遊園地名
- ⑥公園名
- ⑦映画館名
- ⑧温泉
- (4) イベント名
- (5) 検索結果
- (6) 地名
- (7) 鉄道駅名
- (8) 基本的な対話語彙

- ・肯定、否定
- ・問い合わせ
- ・説明、状況通知、確認、……

これらの構成語彙にはデータベースの要素となる固有名詞のみならず、対話上の同義語（はらへった、ごはんたべたい、etc）も含まれる。この各々からここでは1回のワードスポットティングの語彙即ち、目的地をベースとしたレストラン名と関連する料理名（ここではユーザプロファイルも参照する）で100語の大半を構成する。これをもとに『中華街』や『しゅうまい』などが認識語彙としてヒットする。

【0074】[2.4 単語間のネットワーク] 辞書の基本構造は上記の階層表現に準じて定義するが、その他の意味的関係などのネットワーク関係は隨時、ユーザやデータベース供給者から提供される。例えば、ユーザ発話は下記のような属性の組（対話ベクトルと呼ぶ）の集まりである対話データベース（図8）の中で位置づけられる。

【0075】（話題、時間・位置、環境・状況、状態・要求、ユーザ発話、エージェント発話、制御出力）対話ベクトルは無数に存在しうるが、あらゆるベクトル値を取るわけではなく、人間と機械との間の実際的なコミュニケーションの単位として、意味のある有限個のまとまりにクラスタリングできる。そこには単語の意味的な分類、文法的制約、話題の連続性、物理的・常識的制約、事象の連続性などが用いられる。したがって、

(A) あるユーザ発話を構成する単語列に用いられる語彙の範囲

(B) 現在の発話をから次の発話に至る際の語彙の制約は対話ベクトルが張る空間を構成する主要因となる（話題、時間・位置、環境・状況、状態・要求）に大きく影響される。

【0076】そこで、あるユーザ発話における単語が他の単語に対してどういうネットワーク構造になるかは下記の要因で決定する。

- (1) クラスタ辞書間の関係
- (2) 単語間の関係
- (3) 話題間の関係
- (4) 文脈の連続性
- (5) ユーザの特性や状況

#### (6) アプリケーション間

以上の関係に基づいてある単語W1から別の単語W2が活性化され、これを次のユーザ発話に対する認識辞書の語彙に加える。さらに、認識結果に付随する尤度値LFKを高めるように音声認識モジュールのパラメータを調整する。ここで、(1) クラスタ辞書間の関係、(2) 単語間の関係、(3) 話題間の関係、(4) 文脈の連続性、(5) ユーザの特性や状況に関して補足説明する。

【0077】[2.4.1 クラスタ辞書間の関係] 基本的には、上述した[2.3] 辞書の構成で述べた関係に準ずる。

(例) 施設→スキー場→おんたけスキー場

[2.4.2 単語間の意味的関係]

[2.4.2.1 包含関係]

(例) 中華料理→シュウマイ、ラーメン、ギョーザ、…

…

(例) スポーツ→テニス、スキー、スイミング、ジョギング、…

[2.4.2.2 連想関係]

(1) 同一分類のオブジェクトを連想する場合

(例) うどん→麺類+ラーメン

(2) シーンの構成要素を連想する場合

(例) ゲレンデ→スキー→リフト、スノーボード、ゴーグル、…

(例) ゴルフ→ゴルフ場→ホール、キャディ、フェアウェイ、クラブ…

(例) 海辺→海水浴→水着、ビーチパラソル、青い空、白い雲、…

(3) シーンに関連する興味の対象を連想する場合

(例) スキー→ゲレンデ、雪質、リフト…

(例) ゴルフ→天気、経路、費用、スコア、…

(4) 季節から代表的なシーンを連想する場合

(例) 夏→プール、海水浴、かき氷、セミ、クーラー、…

(5) 要求キーワード間に基づく連想

(例) はらへった→レストラン

【2.4.3 話題間の関係】現在の話題に連関した話題のキーワードを活性化することにより、認識語彙を設定することができる。その連関のカテゴリは手段、付随する行動、よくある付帯事象、などがある。

【0078】(例) ショッピング→駐車場（手段）、レストラン（付随する行動）、バーゲン（付帯事象）…

#### [2.4.4 文脈の連続性]

【2.4.4.1 話題の連続性】通常の自然な対話に見られるように、ある話題（たとえばショッピングなど）で閉じた認識語彙の範囲で対話が継続することが考えられる。このような話題の連続性という制約のもとで認識語彙を設定することができる。

【0079】[2.4.4.2 発話-応答の妥当性]

50 車室内に代表される対話環境では、ある発話内容（ユー

ザもしくはエージェントによる)は、(呼びかけ、申告、通知、教示、解説、指示、依頼、警告、督促、問い合わせ)のいずれかに分類できると考えることができる。一方、この発話に対する応答は、(応答、確認、保留、判断、回答、その他応答)に分類できる。この発話と応答の組み合わせを発話対、あるいは対話ユニットと呼ぶ。この対話ユニットに基づいて話題の内容によらず、文脈の論理的な連続性を定義することができる。図9中に「○」で示した部分は対話ユニットとして成立する発話-応答の組み合わせを示す。この対話ユニットをもとにしてエージェントは次のユーザ発話に含まれる認識語彙を予想して設定することができる。

【0080】以下、発話内容と、この発話に対する応答の具体例について説明する。なお、ユーザの発話については「…」で示し、エージェントの発話については「…」で示す。

## (1) 呼びかけ

一般的な意味での呼びかけや挨拶などがこれに含まれる。

## 【0081】(例)

呼びかけ：「おい、XYZ。」

返事：『はい、何ですか。』

## (例)

呼びかけ：『おはようございます、今日はいい天気ですね。』

返事：「ああ、おはよう。」

## (2) 申告

## (例)

申告：「今日は家族とドライブ。」

確認：『御家族とドライブですね。』

## (3) 通知

## (例)

通知：『およそ1km先、渋滞です。』

無応答：「」、又は

確認：「わかった。」

## (例)

通知：『私の名前はXYZです。』

確認：「OK。」、「よろしく。」

## (4) 教示

## (例)

教示：「今、雨が降ってきた。」

確認：『“現在、雨が降っている”というメッセージを確認しました。』

## (5) 解説

## (例)

解説：『操作方法がわからないときはヘルプといってください。』

無応答：「」

## (6) 指示

## (例)

指示：『ユーザパスワードをしゃべってください。』

確認：「わかった。xxxxxx」

## (7) 依頼

## (例)

依頼：『そろそろガソリンが少なくなってきた。次の交差点のガソリンスタンドで給油していただけませんか？』

保留：「いや、あとにしよう。」

## (8) 警告

## 10 (例)

警告：『10km先、○○トンネルで事故発生。次のインターで降りてください。』

判断：「わかった、そうしよう。」

## (9) 督促

## (例)

督促：『ユーザ名がまだ登録されていません。すぐに登録してください。』

確認：「わかった。」

## (10) 問い合わせ

## 20 問い合わせには次の4種類がある。

## 【0082】①合意要請

(例)：『御出にならないので電話接続を中止します。よろしいですか？』

## ②選択要請

(例) 問い合わせ：『Aですか？ Bですか？』

回答：「Aです。」

## ③問い合わせ

場所、時間、情報など特定データの問い合わせをするユニットである。

## 30 【0083】(例) 問い合わせ：『○○○○スキー場の積雪情報はどうなっている？』

(例) 問い合わせ：『これからどちらへいかれますか？』

## ④話題の確認

文脈から外れた突然の話題遷移が発生したことをユーザに確認する。

【0084】(例)：「12時に岡崎にいく。」「ねむい。」

問い合わせ：『“ねむい”と聞こえましたけど、岡崎の40 話はどうなりましたか？』

【2. 4. 4. 3 対話ユニット間の接続性】上記の対話ユニット内の呼応関係のみならず、対話ユニット間の接続(話題の遷移や呼び出し、終了を含む)の妥当性に関する制約も認識語彙の設定において考慮することができる。

【0085】【2. 4. 5 ユーザの特性や状況】ユーザ発話に付随するユーザの環境・状況・要求・状態、ユーザプロファイルに基づいて次のユーザ発話に対応できる認識辞書を設定する。この場合、必ずしも上述の文脈50 の連続性が保たれるとは限らない。

## 【0086】(1)自然な要求推定

例えば本願出願人が特願平10-184840号にて提案したような要求推定装置に基づくと共に、図2に示す要求・状態推定用データ記憶部36に記憶された要求・状態推定用データを参照し、ユーザの環境・状況・要求・状態、ユーザプロファイルから次の認識語彙を限定する。

## 【0087】(2)突然の運転状況の変化

不連続的に発生する予測不可能な緊急事態や警告の対象となる事態に際して、文脈の連続性を一時停止し、事態に必要な話題を割り込ませるべく認識語彙辞書を設定する。

## 【0088】(例)

エージェント：これからどうなさいますか？』

ユーザ：「○○駅前で買い物」

エージェント：『到着時刻は11時ごろです。駐車場はどこにしますか？』

(先行車が急停止したので急ブレーキをかけた)

ユーザ：「あー、危なかった。」

エージェント：『危なかったですね、安全運転にごろがけてください。お疲れならば休みますか？』

## (3)システムの機能移行

機能が切り替わったときにシステムが発話し、必要な対話を開始するべく認識語彙を設定する。

【0089】[2・5 多段階処理の具体例] 多段階処理の具体例を、図10、11のフローチャートを参照して説明する。ここでは、まず使用頻度の高いローカル情報を優先するかどうかを判断して(S210)、認識処理を2つにわける。なお、分岐条件はこれ以外にも考えられ、対話戦略に依存する。

【0090】ローカル優先の場合は(S210: YES)、代表的な場所、施設名、要求キーワード、要求関連属性でN語の辞書を構成し、ワードスポットティングをかけ(S220)、要求キーワードの尤度を構文評価で補正する(S230)。そして、キーワード属性の重み付けによって尤度を補正し(S240)、単語列の順序付け(S250)を行う。その後、認識完了した音声区間を次回の認識対象から外し(S360)、全音声区間を認識完了していなければ(S370: NO)、不足する属性の語彙を追加して辞書更新の準備をしてから(S380)、S210へ戻る。

【0091】なお、この場合はS220～S250がローカル優先の場合の処理であり、この処理は「多段階処理」ではない。そして、このローカル優先の場合の処理によれば、1回の認識ですべての音声区間を処理可能であり、認識時間も少ないが、検索対象が例えば100語に収まるように限られるため、日常的な要求から外れた発話の場合は誤認識になる確率が高くなる。

【0092】一方、ローカル優先でない場合、すなわち要求を優先する場合は(S210: NO)、まず、1回

目の認識か否かを判断し(S260)、1回目の認識であれば(S260: YES)、認識属性を要求キーワードに設定して(S270)、S280へ移行する。2回目以降の認識であれば(S260: NO)、S270の処理は実行せずS280へ移行する。S280では、認識属性のキーワードを多く含む単語セットを構成し、その構成された単語セットを用いてワードスポットティングを実行する(S290)。

【0093】その後、要求キーワードの尤度が所定値(ここでは一例として0.6とする。)以上かどうかを判断し(S300)、要求キーワードが0.6以上の尤度を持っていれば(S300: YES)、S310へ移行する。S310では、同一属性の単語が複数ある場合は構文位置を優先する。つまり、これで要求キーワードに基づく要求が確定し、続くS320においては、ヒットしたキーワードに対応する属性を次回の認識語彙に設定する。その後は、認識完了した音声区間を次回の認識対象から外し(S360)、全音声区間を認識完了していなければ(S370: NO)、不足する属性の語彙を追加して辞書更新の準備をしてから(S380)、S210へ戻る。

【0094】一方、要求キーワードの尤度が0.6未満の場合は(S300: NO)、S330へ移行し、他の属性で尤度が0.6以上の単語があるか否かを判断する。そして、尤度が0.6以上の単語があれば(S330: YES)、その中で最も尤度が高い単語を認識属性とし、その認識属性のキーワードを多く含む単語セットを構成してワードスポットティングを実行する(S340)。しかし、尤度が0.6以上の単語がなければ(S330: NO)、場所、施設名、要求関連属性でN語の辞書を構成し、ワードスポットティングを実行する(S350)。S340、S350の処理後は、認識完了した音声区間を次回の認識対象から外し(S360)、上述したとおり、全音声区間を認識完了していなければ(S370: NO)、不足する属性の語彙を追加して辞書更新の準備をしてから(S380)、S210へ戻る。

【0095】一方、全音声区間を認識完了していれば(S370: YES)、単語列の推定をし(図11のS390)、エコーバックして(S400)、不足条件を問い合わせる発話を(例えは、違う、そうじゃない、など)かどうかを判断し、否定的発話でなければ(S420: NO)、不足する属性の語彙を追加して辞書更新の準備をしてから(S430)、図10のS210へ戻る。一方、否定的発話であれば(S420: YES)、S440～S470の誤認識対応処理を実行する。

【0096】具体的には、S440にてモード設定に基づいて分岐し、連続単語認識に切り替えてユーザに発話方法を指示するか(S450)、ヘルプモードとして要求キーワードを入れるようユーザに指示する(S46

0)、漸進的階層探索として認識可能な語彙を表示するか (S 470)、のいずれかを実行する。その後は、不足する属性の語彙を追加して辞書更新の準備をしてから (S 480)、図 10 の S 210 へ戻る。

【0097】このような処理を、例えば「えーと、あんじょうでらーめんをくいたい」といったユーザ発話音声に対して実行した場合について説明する。まず、音声区間から要求キーワード (くいたい、など) を優先的に検出する。このために、認識語彙も要求キーワードを多く含む (要求の種類や同一要求に対する同義語・類義語・連想語を充実させる) 構成にする。しかし、この要求優先の場合、1回のワードスポットティングでは認識が完了しない。したがって、図 6 に示したように要求キーワードが確定 (食事関連) した後に次の施設名や要求対象をそれぞれ1回のワードスポットティングに対応させて認識を繰り返し、すべての音声区間を認識終了した時点で単語列推定を駆動させて単語列認識結果を絞り込む。なお、この単語列認識結果の絞り込みの詳細については、例えば特願平 11-20349 号を参照されたい。

【0098】これをもとにシステムは「あんじょうでらーめんをたべるんですね」といったエコーバックの発話をを行い (S 400)、論理的発話戦略の駆動により不足条件を問い合わせる発話 (S 410) として「あんじょうには3けんらーめんやがあります。みせはどこにしますか?」と発話する。これに対してユーザは、まったく異なる発話「〇〇でれいぞうこをくいたい」と発していたとすると、上述のエコーバック結果に対しては否定的応答 (「ぜんぜん違う」など) を発話する (S 420: YES)。したがってその場合は誤認識対応処理 (S 440~S 470) が駆動される。

【0099】一方、否定的応答がなされない場合 (S 420: NO)、ユーザとの間で正しく対話がなされないとエージェントは解釈し、次にユーザが発話すると期待されるキーワード群 (いっけんめ)、「すしやにかえて」など) を認識語彙辞書に設定して (S 430)、ユーザの発話を待つ。

【0100】[3. 誤認識対応処理について] 音声認識部から出力される認識語彙の尤度判定と語彙限定、ユーザからのヘルプ要求、画面表示の支援に基づく入力の簡便化 (スイッチ操作の併用)、不足条件の判定、サービス内容の有機的連関付けなどに基づきヘルプ、選択肢の提示、初期化、問い合わせを駆動する。

【0101】[3. 1 誤認識の検出] エージェントがユーザ発話を認識できていない状況は誤認識であり、以下のケースがある。

- ①どの認識辞書にもない言葉をユーザが発話
- ②他の辞書にはあるが、現在使っている辞書にない言葉をユーザが発話
- ③ユーザ以外の話者の発話内容に応答し、ユーザの意図に反してモード移行

このような状況は、エージェントの応答に対してユーザが「ちがう」、「わかつてないね」、「全然だめ」といった発話をすることで検出される。この場合、下記のいずれかの対話内容をユーザの状況に応じて選択する。

【0102】(1) 音声メニュー モード

(2) 選択肢を出す

なお、例えば「ちがう」などの発話がユーザから K 回以上繰り返されたときは初期状態に戻る。K は例えば 5 とする。

10 【0103】[3. 1. 1 期待外の応答であることの検出]

認識語に付随する尤度のしきい値処理に基づき、尤度が低い場合は音声認識の信頼度が低いため認識語彙以外の発話がなされたとして、【誤認識の可能性有り】と判定する。

【0104】[3. 2 話題転換の検出] ユーザ発話がエージェントの期待から外れているかどうかはそれまでの文脈に沿っているかどうか、即ちある話題の中で予想される発話かどうかで判定する。予想される発話は

20 「2. 4 単語間のネットワーク」で示した単語間の関係をもとに導出され、これに対応する語彙を認識辞書に設定する。これを便宜上ケース A と呼ぶ。それ以外の場合は、予想されない発話であり、それを構成する単語は下記のように分類できる。

【0105】(B 1) 認識語彙辞書に登録されていない単語

(B 2) 認識語彙辞書に登録されているが、違う話題の単語

(B 2 a) 現在の認識語彙範囲に含まれている単語

30 (B 2 b) 現在の認識語彙範囲に含まれていない単語  
このうち、(B 1) と (B 2 b) は通常のワードスポットティングでは認識され得ないため、不用語とみなされるかあるいは信号処理的に近いとみなされる他の認識可能語彙に置き換えられて出力される。これらは後述の誤認識処理で対応する。

【0106】一方、(A) と (B 2 a) については以下の 3 つの処理形態で対応する。

(1) 文脈優先処理 [(A) の場合]

40 出力された複数の認識候補 (ラティス) 間に尤度の差異があり認められない (分散が小さい) とき、文脈に沿った認識候補が優先して選ばれる。

【0107】(2) 突然の話題の遷移 (話題転換)

[(B 2 a) の場合]

出力した認識候補が 1 個であり、ある一定しきい値以上の尤度を持つ場合、この話題に突然遷移したことを認める。

(3) 突然の話題の遷移の確認 [(B 2 a) の場合]

出力した認識候補が 1 個であるが、ある一定しきい値以上には満たない尤度を持つ場合、この話題に突然遷移したのかどうかを確認するため、ユーザに問い合わせる。

【0108】 [3. 2. 1 話題転換の検出] 現在の話題とは異なる要求キーワードが認識された場合には話題が転換した可能性があるとみなす。

【3. 2. 2 話題転換の確認】 上記に基づき、システムはユーザに対して話題が転換したことを確認する問い合わせの発話を生成する。

【0109】 (例)

エージェント『□□駅前には3件のラーメン屋があります。』

ユーザ 「東京の○○さんに電話をかける」

エージェント『電話をかけますか？』

ユーザ 「うん。東京の○○さんに」

(例)

エージェント『○○駅前には3件のラーメン屋があります。』

..... (T(n)=食事)

ユーザ 「あっ、電話かけなきや。」

..... (『電話』が認識され

ればTnew=電話だが話題を更新しない)

エージェント『どのラーメン屋にしますか？』

..... (T(n+1) = T(n)=食事)

ユーザ 「えーと、□□」

エージェント『□□へのルートを表示します。』

なお、上述した話題転換と、ここで説明した文脈優先は相反する応答だが、例えば、その条件判断はU(n)の尤度情報LFKやU(n)で示される話題Tnewの連続出現回数Nnewなどを用いる。即ち、 $N_{new} > 2$  且つ LFK  $> 0.4$  ならば話題はT(n-1)からT(n)=Tnewに移行し、それ以外の場合はU(n)から得られたTnewは棄却し、T(n)=T(n-1)とする、といった条件分岐を用いる。

【0112】 [3. 4 ヘルプモード] 起こりやすい誤認識の例、代表的な要求キーワード、などの選択肢を表示したり音声で発話する。

(例)

・電話をかける場合は、「でんわをかける」で電話番号画面を表示し、相手先の電話番号を入力してください。登録されている相手先（例えば○○さん）の場合は「でんわをかける、○○」でも結構です。

【0113】 ・次のどれですか？もういちど発話してください。 {食事（レストラン、ごはん、おなかすいた）、スケジュール帳、アドレス帳（住所録、電話帳）、地図（経路案内、ドライブアシスタント）}

・かっこ内の同義語でもう一度発話してください。

【0114】 ・地図表示の場合は、経路案内から道路図を選択したほうが確実です。

・目的地は市町村をつけて発話した方が正確です。（例 かりや→かりやし）

【3. 5 選択肢の提示】 選択肢提示は既に説明した漸進的階層探索に相当する。

【0115】 [3. 6 通常の発話戦略] 認識結果（認

エージェント『東京の○○さんに電話をかけます』

【3. 2. 3 話題転換の発話】 話題が転換したことを宣言する発話を生成する（上記例参照）。

【0110】 [3. 3 文脈優先の発話] それまでの話題T(n) (n:発話対の通し番号) が継続していると仮定し、次のエージェント発話もその話題に基づいて生成する。従って、今回のユーザ発話の解釈結果U(n)が話題T(n)に関連しない語彙であっても、それに即応せずに、話題T(n)からの文脈に沿って発話内容に限定処理を加える。

【0111】

10

エージェント『どのラーメン屋にしますか？』

..... (T(n+1) = T(n)=食事)

ユーザ 「えーと、□□」

エージェント『□□へのルートを表示します。』

識語、尤度) をもとに誤認識が検出されなかった場合は通常の発話戦略が適用される。この通常の発話戦略の内容は本発明の主眼とするところではないので、ここでは詳細については言及しない。

【0116】 [3. 7 対話管理の適応化]

【3. 7. 1 環境・状況への適応】 時間（季節、日時、時刻【朝昼夜】）、空間（自車位置、地域【都道府県、市町村】）、環境（道路環境【高速、一般道、トンネルなど】、道路状態【路面凍結、滑り易い、他】、交通環境【高速道路、速度制限など】、地理環境【海が近い、山の中、街中、駅前、など】）、車外状況（天候、交通状況【渋滞など】、車外周辺状況【追い越し車あり、など】）、車内状況（運転状態、乗車状態、移動目的、話題）などに応じて、話題の選定や対話管理、適切なメッセージの伝達などを行うことができる。なお、これらは表示系にも反映される。

【0117】 [3. 7. 2 スケジュールへの適応]

40 (1) ドライブスケジュール

ユーザの設定した目的地や経由地に基づき、ドライブスケジュールを作成し、各イベントの意味（食事、ショッピング、観光など）と場所・時間に基づいて、話題の決定や対話管理、さらにはドライブスケジュールの空白部分について提案することができる。

【0118】 (2) 個人スケジュール

PDAやPC用のインターフェースを介してPDAやPC上の個人スケジュールデータをダウンロードし、それに基づいて話題の決定や対話管理、さらには個人スケジュールの空白部分について提案することができる。

50

【0119】 [3. 7. 3] ユーザが応答の意味を理解できないとき] ユーザが「どういう意味」、「よくわからん」といった発話をしたとき、ユーザはエージェント応答の意味を理解できていないと判断し、次のいずれかの処理を行う。

【0120】 (1) ヘルプ機能

(2) メニュー選択

(3) 音声メニュー

【3. 7. 4】 エージェントがユーザ要求に対応できないとき] ユーザ発話を正常に認識できても、以下のような場合ではエージェントがユーザの要求に対応できないことがある。これらについてはその状況を音声メッセージで伝達する。

【0121】 ①検索結果がない (該当データベースがない)

②該当する機器がない、故障中あるいは準備されていない (電話が接続されていない、など)

③操作対象となる機器の動作範囲を超えた制御命令が出されたとき

【3. 8 音声メニュー モード】 例えば本願出願人が特願平10-177667号にて提案したような機器操作モードに移行する。各操作モードにおいて必要なコマンドは必ず優先的に認識語彙辞書に加え、認識の尤度を高める。

【0122】 [3. 9 問い返し] 問い返しは下記のようにいくつかの場合が考えられ、問い合わせ場合にはその内容に沿った語彙に限定した認識辞書が設定される。

(1) エージェントが正しく話題を認識しているにもかかわらず、それに沿ったユーザ発話がなされなかった場合、話題確認のための問い合わせを発することにより、それ以後の誤認識を回避する。

【0123】 (例)

エージェント『□□駅前には3件のラーメン屋があります。』

ユーザ 「東京の○○さんに電話をかける」

エージェント『電話をかけますか？』

(2) エージェントが話題を間違えて認識し、ちがう話題のところで不足条件を問い合わせた場合、ユーザからの問い合わせ又は否定的発言 (「ちがうよ」や「そうじゃない」など) により、正しい話題が何であるかを問い合わせ。これにより、それ以後の誤認識に歯止めをかける。

【0124】 (例)

ユーザ 「東京の○○さんに電話をかける。」

エージェント『□□駅付近のレストランを表示します。』

どれにしますか？』

ユーザ 「そうじゃない、電話をかける」

エージェント『電話ですか、食事ですか？』

ユーザ 「でんわ」

これは、すでにレストラン検索依頼が出ており、その検索をしている最中にユーザが電話をかけるという割り込

み要求を行った場合である。2つの用件 (電話か食事か) の優先順位を問い合わせた例と言える。

【0125】 (3) エージェントが正しく話題を認識し、文脈上必要な条件が不足していることを確認した上で問い合わせ。これにより、文脈に沿った正しい対話が展開され、問題解決を促進する。

(例1)

ユーザ 「東京の○○さんに電話をかける。」

エージェント『東京の○○さんに電話をかけるんですね。それは○○◇◇さんですか、それとも○○△△さんですか？』

ユーザ 「○○◇◇さん」

エージェント『○○◇◇さんに電話をかけます』

(例2)

ユーザ 「東京の○○さんに電話をかける」

エージェント『□□駅付近のレストランを表示します。どれにしますか？』

ユーザ 「あとにして。先に電話をかける。」

エージェント『東京の○○◇◇さんに電話をかけます。よろしいですね。』

ユーザ 「うん、そうして」

ここで説明した (例1) の場合は、エージェントが正しく認識し、電話をかけるという問題解決に向けて不足している条件を問い合わせた例である。

【0126】 また、(例2) の場合には、ユーザが電話をかける意図がそれまでの対話の経過やスケジュールなどでわかっているならば、その電話の用件 (○○さんと食事する) で必要な情報を事前に検索することが可能になる。つまり、気を回して自動的にレストラン検索を行い、問い合わせた例と言える。

【0127】 [3. 10 初期化] ユーザからの問い合わせがK回 (K>N) を超える場合、初期状態あるいはトップのメニュー画面に戻る。以上説明したように、本実施形態の制御システムによれば、漸進的階層探索、多段階処理、期待外れ対応処理、誤認識対応処理を行うが、それぞれの処理によって以下の効果がある。

【0128】 まず、漸進的階層探索の場合には、上述した「岡崎で食事したいなあ、インド料理がいいね」という発話を例に取れば、ワードスポットティングによる音声認識手法の出力結果が「岡崎、食事、インド料理」という単語列であった場合、従来の音声認識手法では、これら3つが揃った段階で認識に対応するシステム側の処理が開示されているのに対し、本手法によれば、「岡崎」が入力された時点で、次に利用者に入力を期待する発話語彙を即座に提示できる。そのため、利用者はとまどうことなく発話できるようになる。これによって、誤認識の原因となる認識辞書外の語彙を利用者が発話してしまうことを未然に防止できる。

【0129】 また、多段階処理の場合には、単語列を構成する要求キーワードなどを軸として、単語列を構成す

る単語間の意味的な制約を利用して認識辞書を動的且つ小規模に構成することにより、適切な認識を行う。これによって、利用者の意図しない誤認識が発生しないようにできる。

【0130】また、期待外時対応処理においては、対話中における利用者からの発話が文脈に沿った期待通りの内容ではない場合に、話題が転換されたのかどうかを確認するための問い合わせを行う話題転換確認処理、話題が転換されたことを宣言する話題転換宣言処理、それまでの話題が継続していると仮定して文脈に沿った対応を行う文脈優先対応処理のいずれかを行う。一方、誤認識対応処理においては、対話中における利用者からの発話が所定の否定的内容であった場合に、正しい話題が何であるかを確認するための問い合わせ返し処理や、要求する内容に対応したキーワードを含める指示を利用者が視覚又は聴覚にて認識可能なように提示するヘルプモード処理や、初期状態に戻る初期化処理などを実行する。このようにすることで、誤認識に起因して利用者が途方に暮れてしまう、といった不都合を防止することができる。

【0131】さらに、本実施形態の場合には、これら漸進的階層探索、多段階処理、期待外時対応処理、誤認識対応処理が組み合わされているため、さらに効果的である。ところで、本実施形態においては、例えば自動車に搭載されて、ユーザとしての車両の乗員（主に、運転者）と音声にて対話しながら、その車両に搭載された様々な機器を制御するシステムとして説明したが、図1の概念図に示すように、他の信号系認識部にて認識した結果としての単語列候補に対しても同様の処理が可能である。例えば、認識対象者によって入力された手書き文字列を辞書データと比較し、一致度合の高い複数の単語列候補を出力する文字認識装置であってもよい。手書き文字に関してもやはり誤認識が発生し易いため、適正な単語列を推定することは有効性が高い。また、音声認識や文字認識のように、認識装置に入力される時点で直接的に単語列の内容となっているものに限らず、画像認識装置であってもよい。即ち、認識対象を捉えた画像を画面として認識した上で、画面を自然言語化するような認識装置であれば実現できる。具体的には、例えば認識対象者が手話をしている画像から手話パターンを認識し、その手話パターンが表す自然言語的な意味を示す単語列候補を出力するようなものである。手話パターンについても、微妙な指使いによって表す単語が異なるため、手話をする者の個人差などによって、やはり誤認識は発生する。したがって、やはりこの場合も、誤認識が含まれる可能性の高い手話パターンの認識装置においても、上述した各種処理を実行することで、誤認識を防止したり、あるいは誤認識が発生した後の適切な対処を行うことができ、誤認識に起因して利用者が途方に暮れてしまう、といった不都合を防止することができる。

【0132】以上、本発明の一実施形態について説明したが、本発明は、上記実施形態に限定されるものではなく、種々の形態を採り得ることは言うまでもない。例えば、図3の処理概要を示すフローチャートにおいては、多段階処理（S30）が先に行われ、その後の誤認識対応処理の中で漸進的階層探索（S130が相当する）が実行されているが、そのような順番には限定されない。

【0133】また、上記実施形態の場合には、漸進的階層探索、多段階処理、期待外時対応処理、誤認識対応処理が組み合わされていたが、これらは単独で実施しても効果はある。但し、上述したように、組み合わせて実施すればさらに効果的である。その際、4つの処理を全て組み合わせなくてはならないわけではなく、2つ以上の処理の組み合わせであればよい。

#### 【図面の簡単な説明】

【図1】 実施形態の単語列認識装置について機能に着目して概念的に示したブロック図である。

【図2】 単語列認識装置を車載の制御システムに適用した場合の構成を示すブロック図である。

【図3】 システム制御部で実行される処理の概要を示すフローチャートである。

【図4】 漸進的階層探索の具体的な画面遷移例を示す説明図である。

【図5】 漸進的階層探索の具体的な画面遷移例を示す説明図である。

【図6】 多段階処理の概要を示す説明図である。

【図7】 辞書の動的構成の概要を示す説明図である。

【図8】 対話データベース表の構成を示す説明図である。

【図9】 対話ユニットとして成立する発話-応答の組み合わせを示す説明図である。

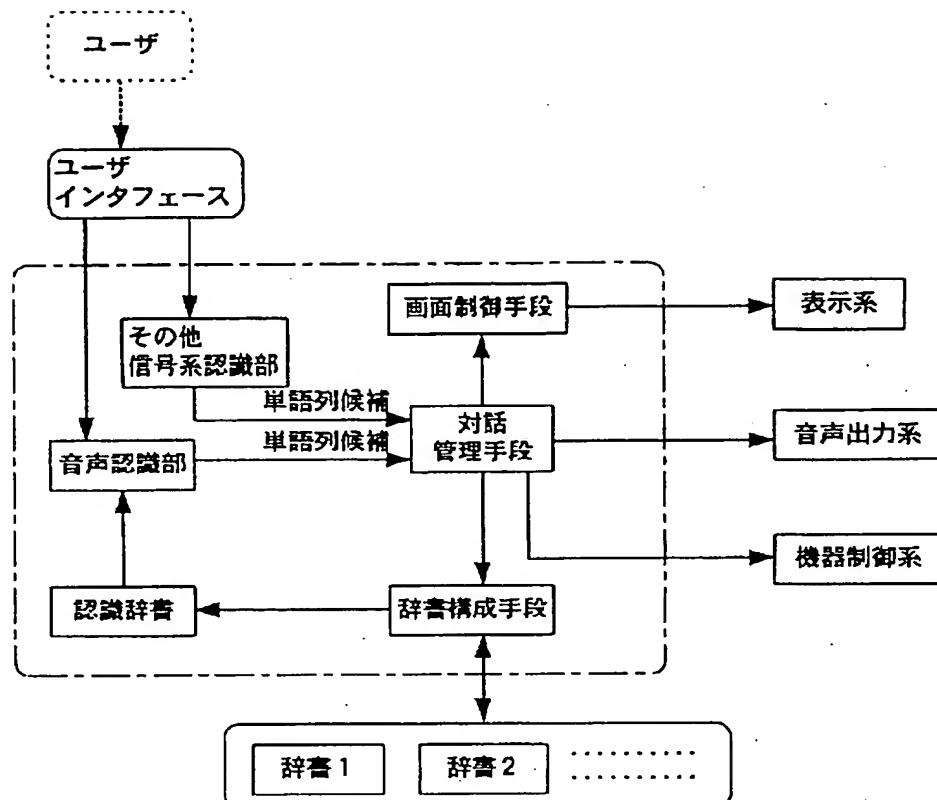
【図10】 多段階処理の前半を示すフローチャートである。

【図11】 多段階処理の後半を示すフローチャートである。

#### 【符号の説明】

1 … 制御装置	3 … 入力装置
5 … マイクロフォン	7 … スピーカ
8 … ディスプレイ	9 … ナビゲーション装置
11 … 表示装置	13 … エアコン装置
15 … オーディオ装置	17 … 通信装置
21 … システム制御部	25 … 音声入力部
27 … 音声合成部	28 … 画面制御部
29 … 機器制御 I / F	31 … インターネットアドレスデータベース
33 … 検索制御部	34 … 認識語彙記憶部
35 … 対話データ記憶部	36 … 要求・状態推定用データ記憶部
37 … ユーザプロファイル記憶部	

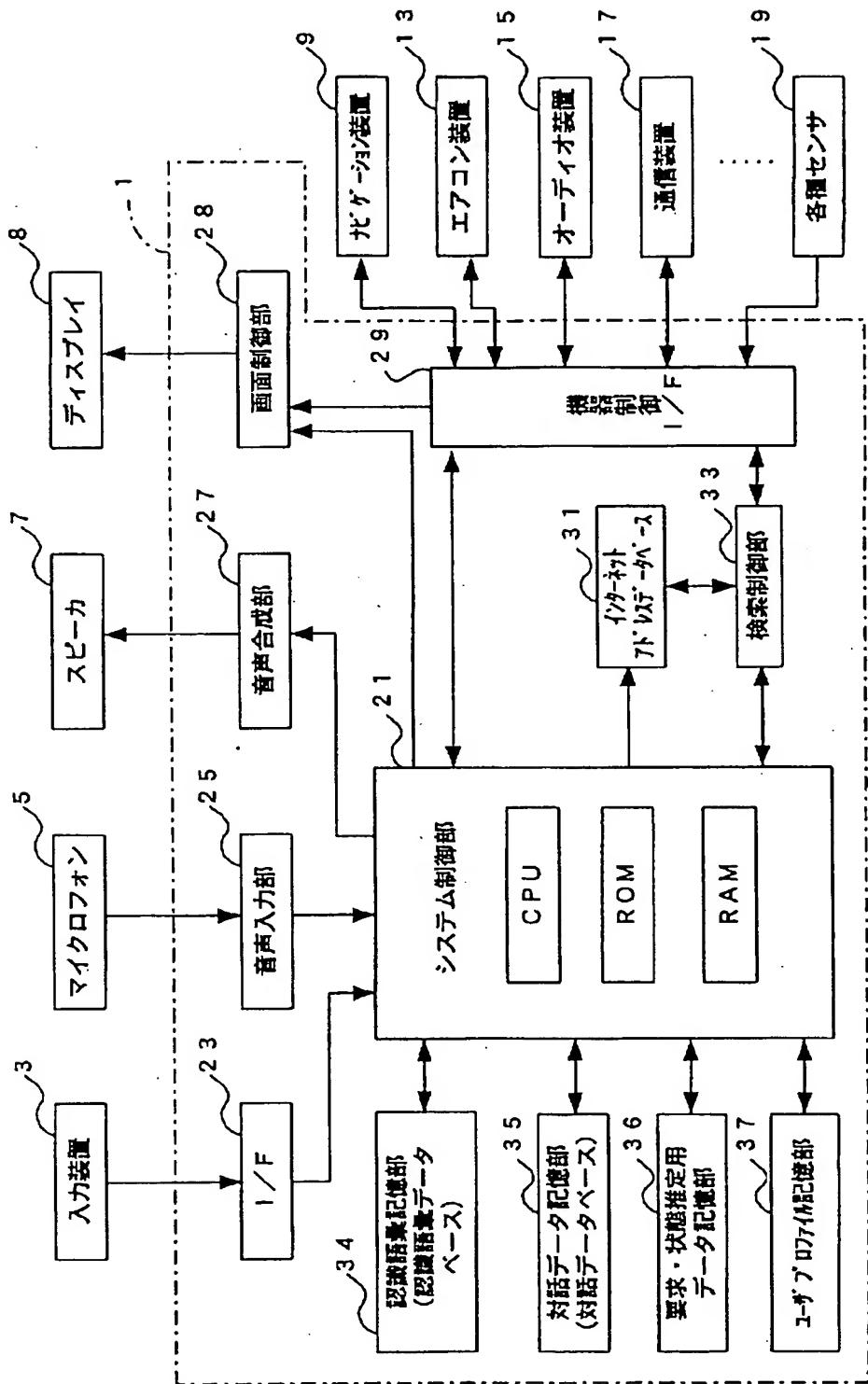
[图 1]



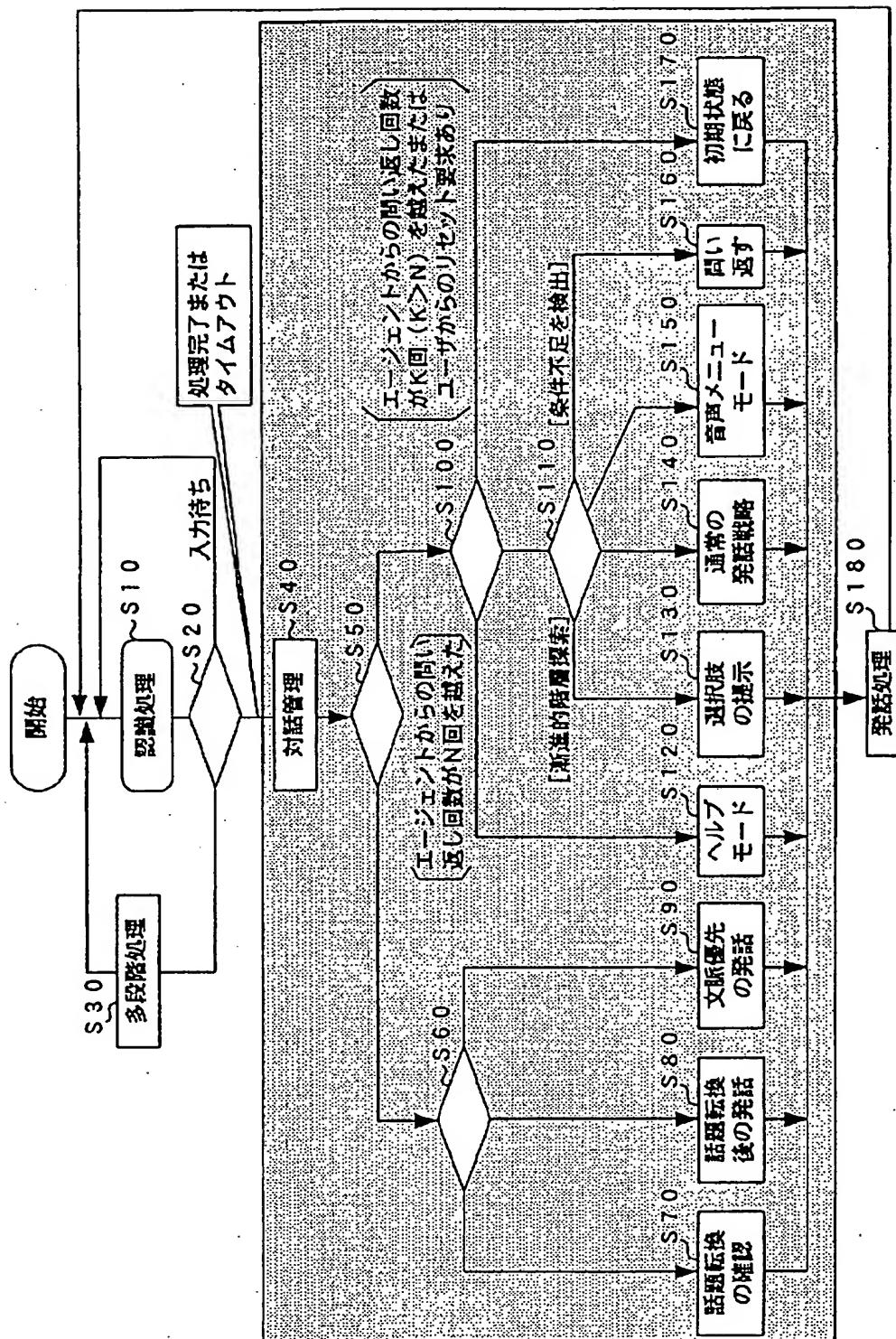
[图 9]

	無応答	確認	保留	判断	回答	その他 応答
呼びかけ	—	○	—	—	○	○
申告	—	○	—	—	—	○
通知	○	○	○	—	—	○
教示	—	○	—	○	—	○
解説	○	○	—	—	—	○
指示	—	○	○	—	—	○
以来	—	○	○	○	○	○
警告	—	○	—	○	—	○
警保	—	○	○	—	○	○
問い合わせ	—	○	○	○	○	○

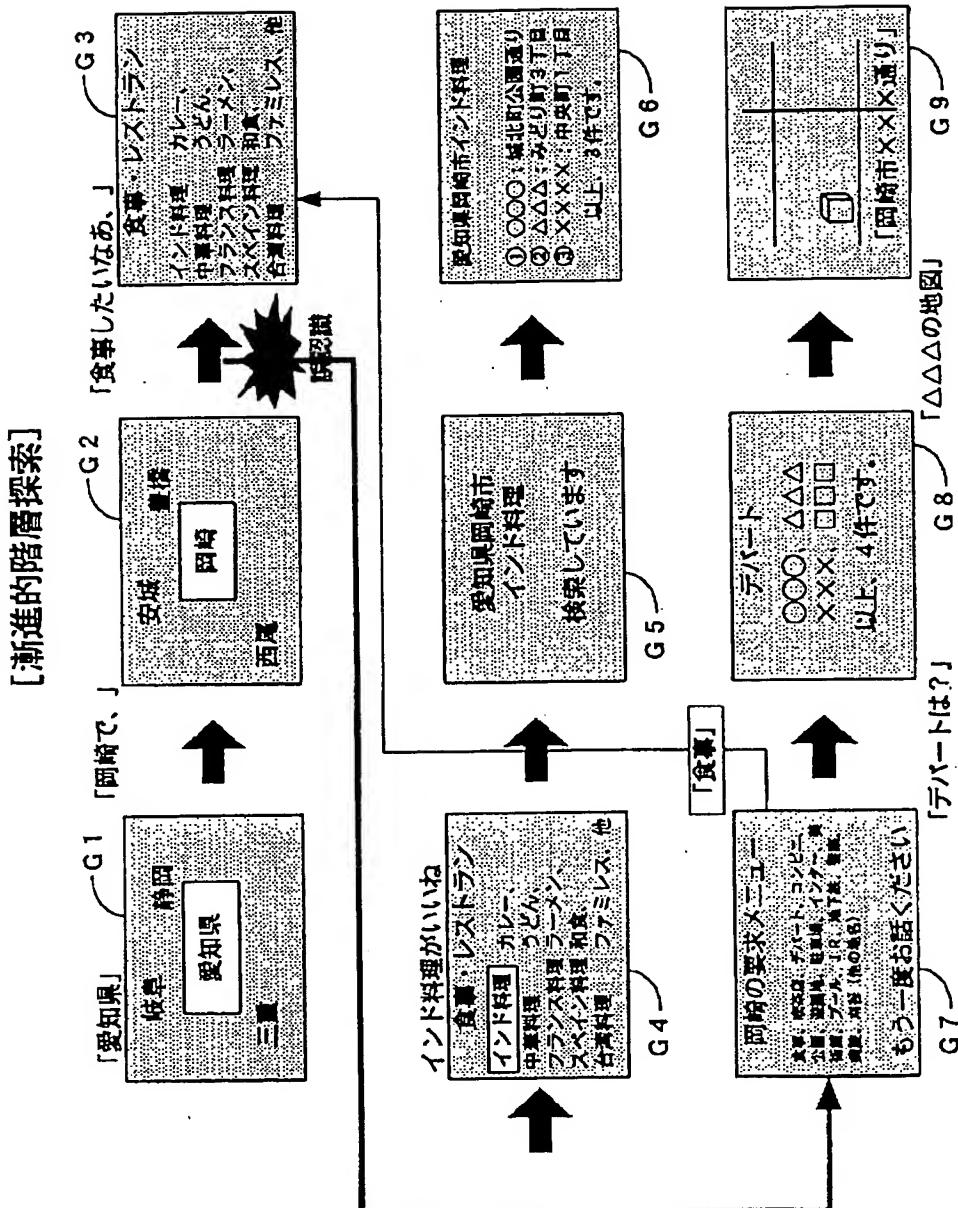
【図2】



【図 3】

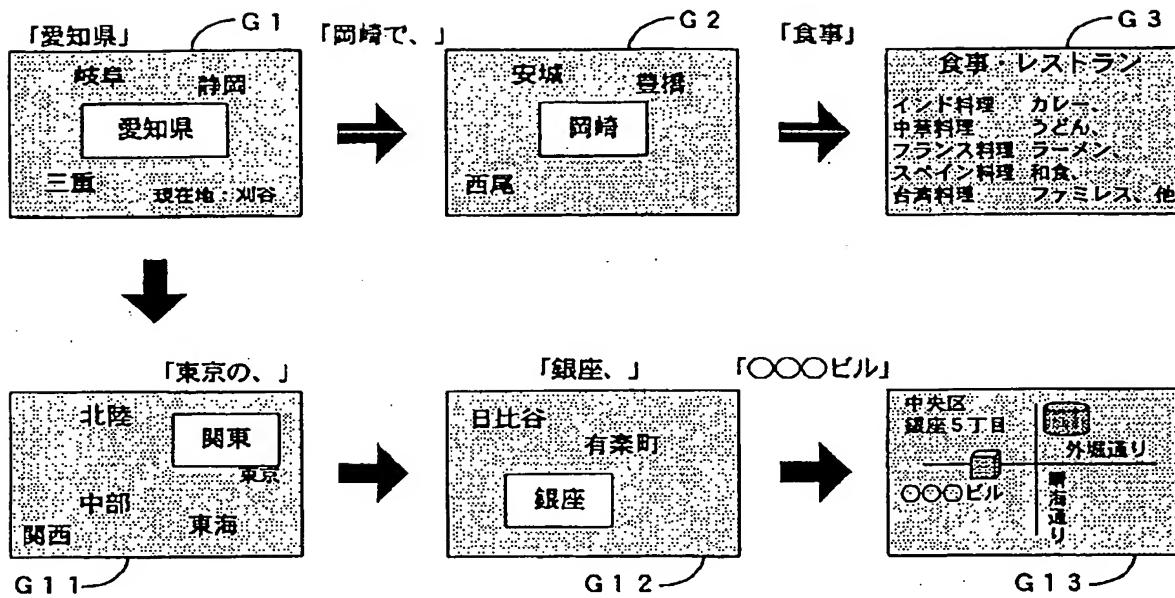


[図4]

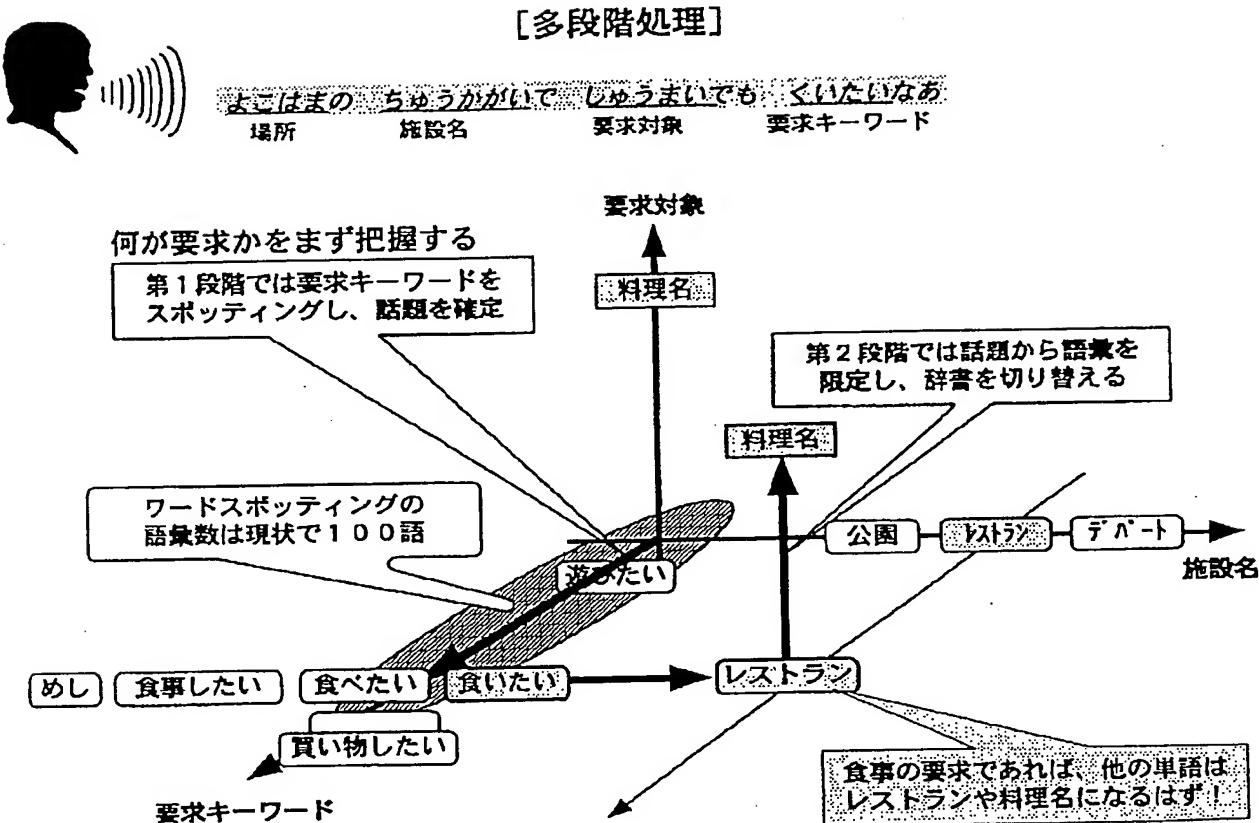


[図 5]

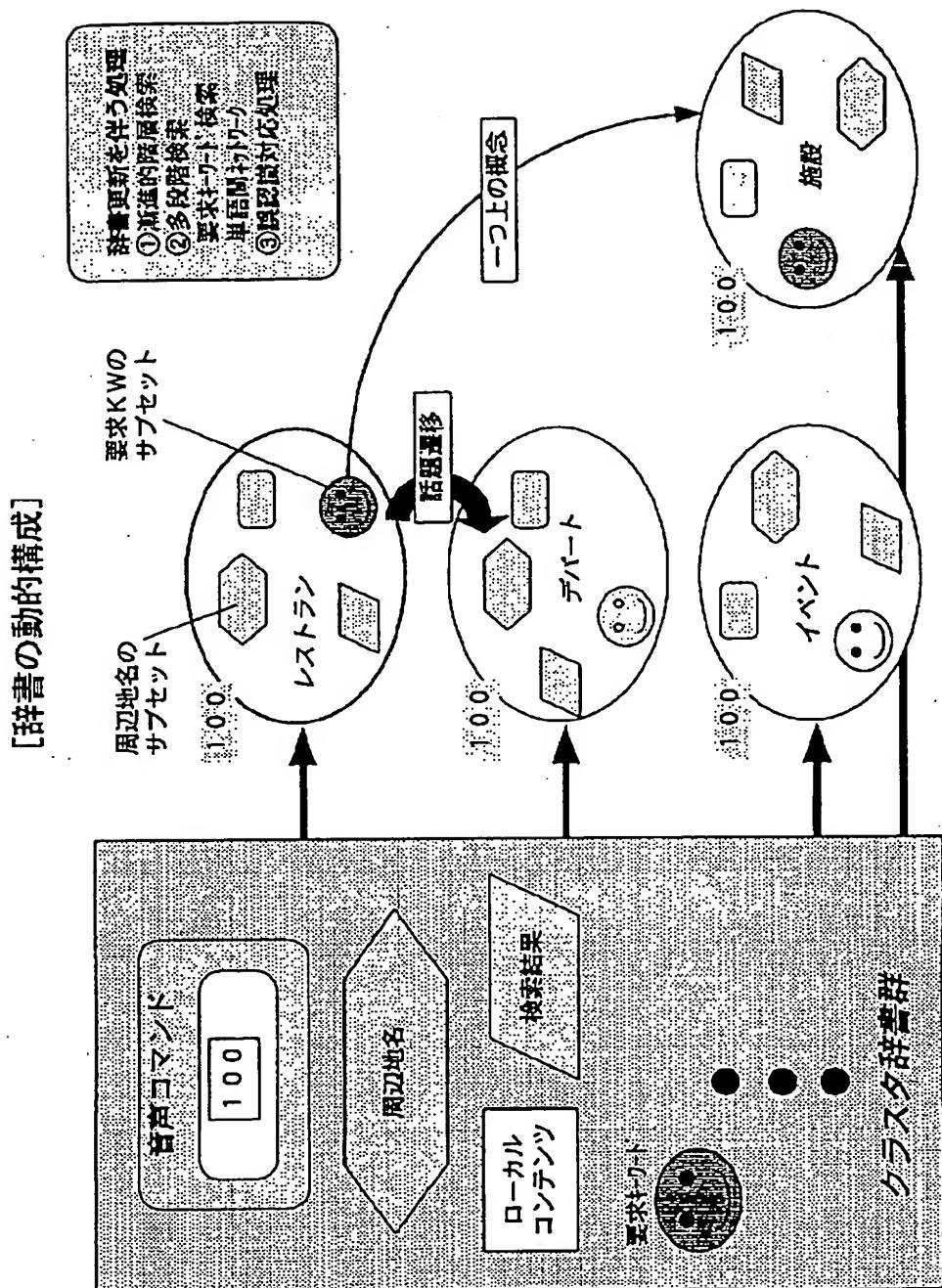
## [漸進的階層探索]



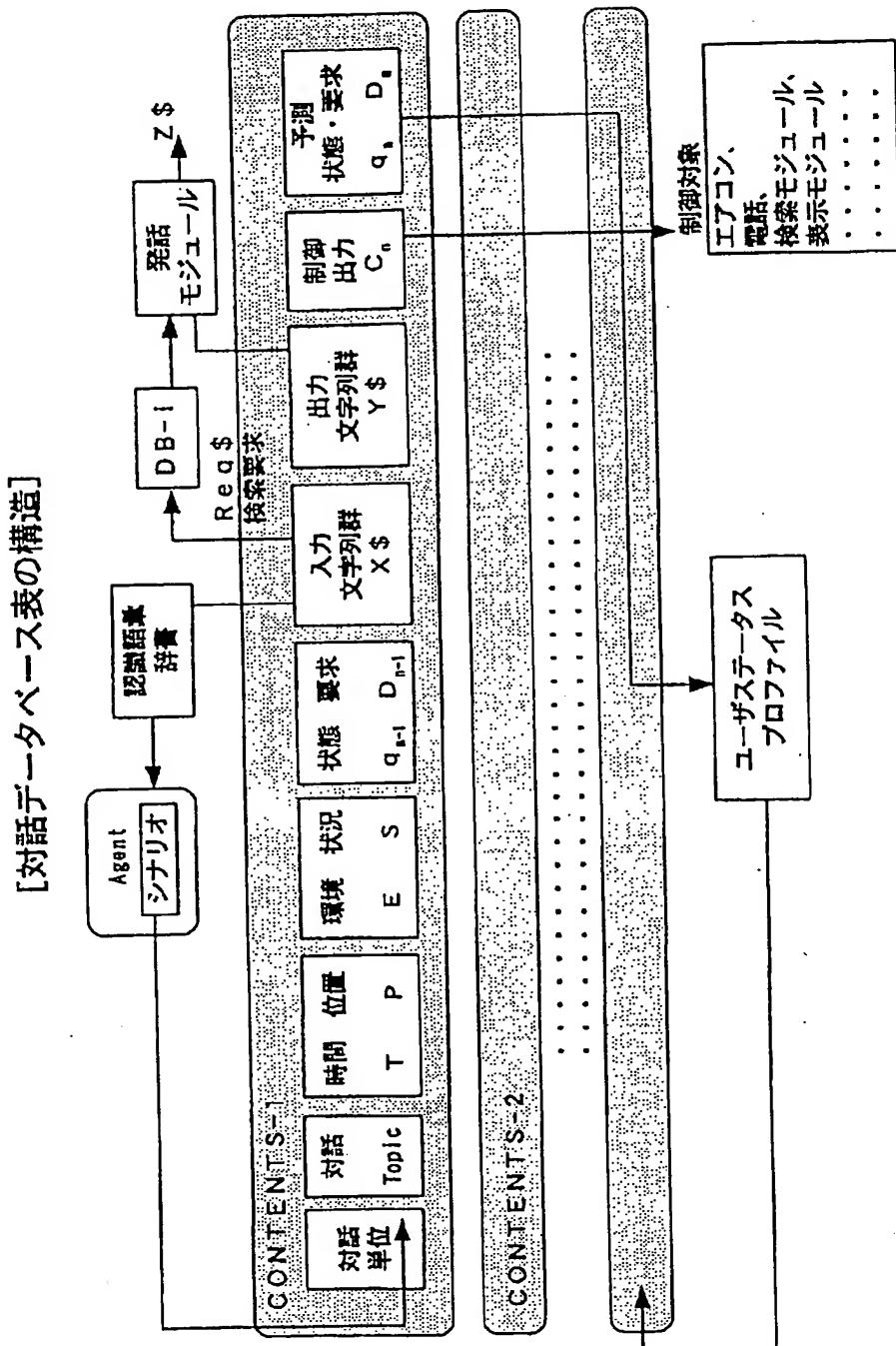
[図 6 ]



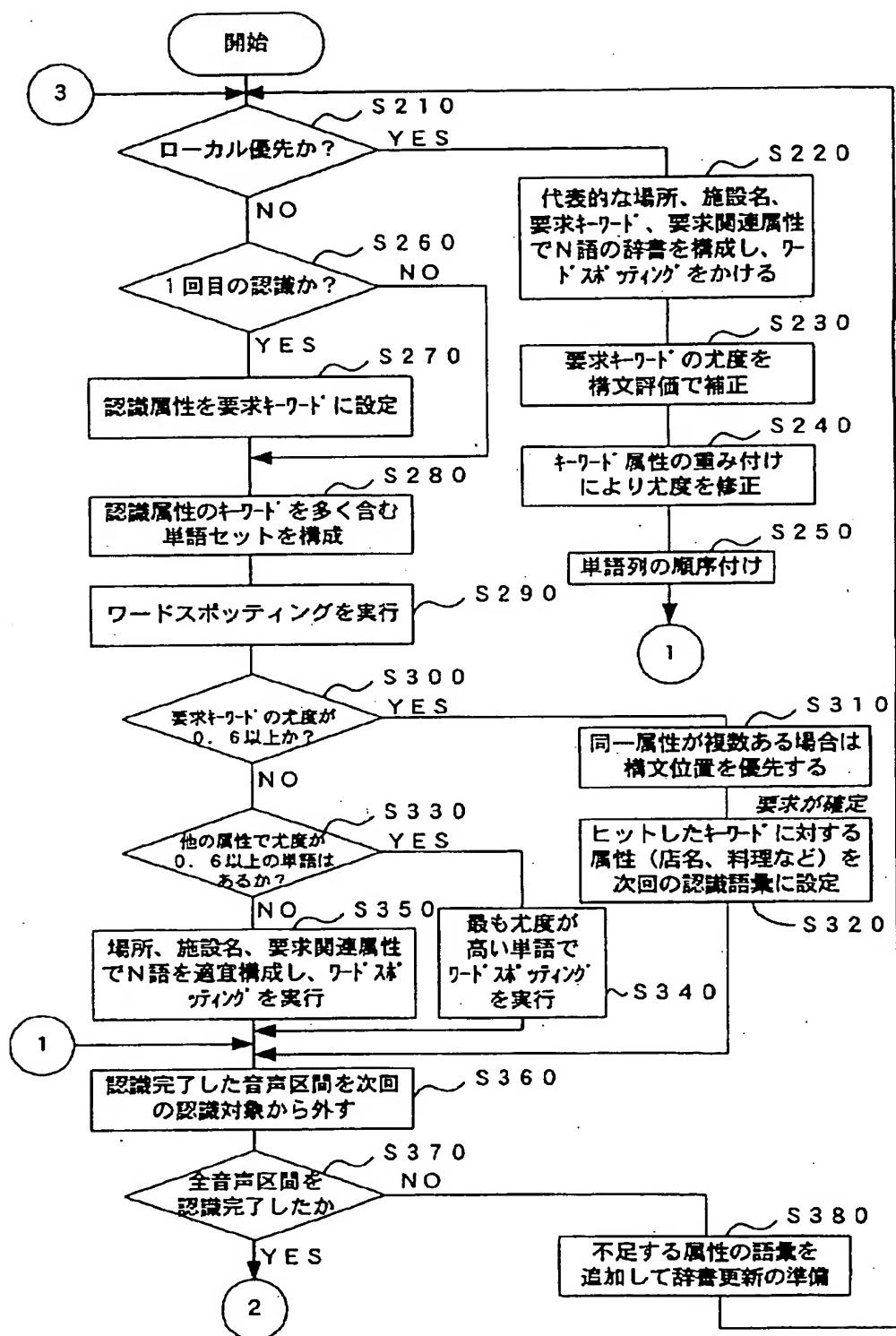
【図 7】



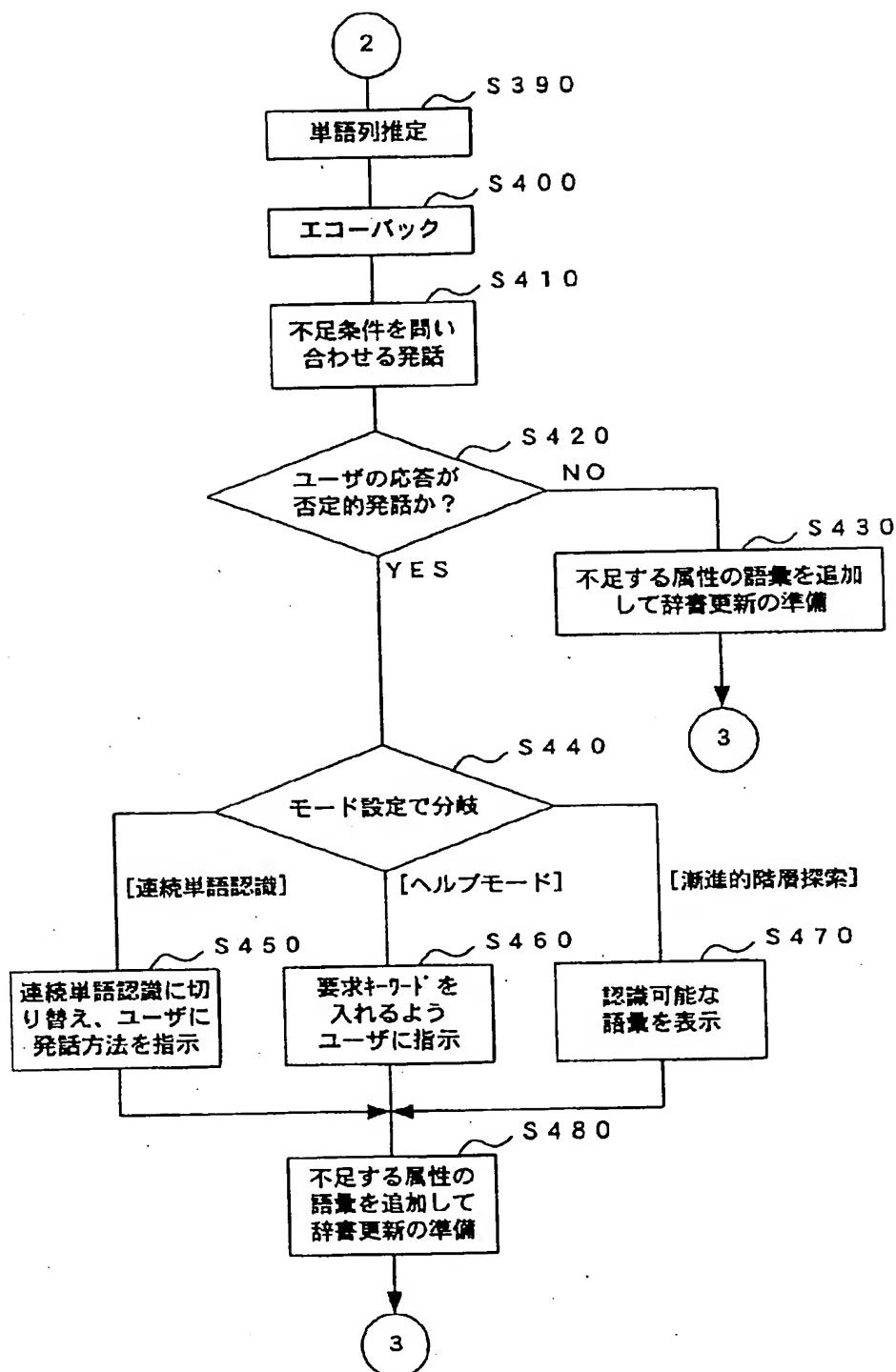
[図 8]



【図10】



【図 11】



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)